The Journal of the

BRITISH INSTITUTION OF RADIO ENGINEERS

FOUNDED 1925 INCORPORATED BY ROYAL CHARTER 1961

"To promote the advancement of radio, electronics and kindred subjects by the exchange of information in these branches of engineering."

VOLUME 23 MAY 1962 NUMBER 5

SONAR SYSTEMS

ALTHOUGH sonar (i.e. *so*und *n*avigation *a*nd *r*anging) is nearly two decades older than radar, and has in many respects maintained a lead in the development of echo-ranging and directional techniques, yet it has somehow never become so well-known, and has sheltered modestly behind a screen of security classification. The Symposium on Sonar Systems which is being held from the 9th to the 12th July this year at the Electrical Engineering Department of the University of Birmingham under the joint auspices of the British Institution of Radio Engineers (through its Electro-Acoustics and Radar and Navigational Aids Groups), the Acoustics Group of the Physical Society, and the University Department, has therefore a rather special significance. It is thought to be the first fully-open conference of such magnitude and scope ever to have been held in the field of Sonar Systems. The response to the preliminary announcement, and the acceptance of 25 papers already, indicate that the need for such a conference is beyond doubt. Nearly half the papers come from overseas, including the U.S.A. and Canada as well as European countries, so there will be a strong international atmosphere at the Symposium.

Sonar—or "asdic" as it has traditionally been called in Britain—is concerned with gathering information about the presence, position and nature of objects by means of sound waves. It most frequently takes the form of an echo-location system used in water, although other forms are possible, and it can be used in air (e.g. as a guidance aid for blind people). Sonar systems occur in nature, being used for example by bats and porpoises. But the emphasis at the Symposium is on man-made underwater systems and the programme has been designed to cover all aspects from "hardware" to pure theory.

The fact that a major field of application is to naval operations is obvious and accounts for the security shroud which has so far shielded the subject. But applications in other, purely civil fields have been growing fast. One has only to turn the pages of journals such as *Deep-Sea Research*, *International Hydrographic Review*, *Journal of the Institute of Navigation* and *Fishing News International* to see how sonar has become an essential instrument in oceanography, geology and geophysics, hydrographic surveying, marine navigation, fishing and marine biological research generally. It seems doubtful if the users of sonar have been fully informed of its modern potentialities, and the lack of open conference facilities may well have hindered even the sonar technologists themselves. Thus the July Symposium should make a large contribution to progress in this field.

D. G. T.

INSTITUTION NOTICES

Symposium of Sonar Systems

The editorial of this issue of the *Journal* discusses the background of the forthcoming Symposium on Sonar Systems, which will be held at the University of Birmingham from 9th to 12th July, 1962. An outline programme and registration form are being circulated with this issue; members wishing to obtain further copies should apply to the Institution at 9 Bedford Square, London, W.C.1. As the accommodation at the Symposium will be limited, early application is advisable, particularly by members overseas.

Radar Display Symposium

The Radar and Navigational Aids Group Committee is planning to hold a one-day symposium of papers on the general theme of "Processing and Display of Radar Data". The meeting will take place in London during March 1963 and offers of papers on all aspects of the subject are invited. Members able to contribute should write as soon as possible giving full details of the proposed paper (provisional title, synopsis, estimated length) to the Secretary, Radar and Navigational Aids Group Committee, 9 Bedford Square, London, W.C.1.

Radio and Electronics Research in Great Britain

Under the above title the Institution's Research Committee is submitting a report to the Council for possible publication.

There is currently much discussion on the possibility of promoting a new research association devoted to radio and electronics, or setting up a branch within an existing research association for this purpose. Under the second category come the various comments which have been made recently regarding the activities of the British Electrical and Allied Industries Research Association—generally referred to as E.R.A.

At a recent E.R.A. Luncheon Lord Fleck stressed the advantages of blending fundamental research with investigation of practical problems. Sir Christopher Hinton took up this point and suggested that a research association undertook three types of work: fundamental, investigational and semi-technical development, with the major portion lying in the middle category. He believed duplication of research to be of little danger in fundamental work where free exchange of information existed. In other more technical fields there was unlikely to be exact duplication and commercial competition would be a spur to these investigations. Sir Christopher stressed the importance of research associations being close to centres of development, but he considered projects of a long-term nature to be desirable as they obviated the need for day-to-day contact with commercial design engineers.

Quebec Engineers Act (1920)

The February 1962 issue of the *Bulletin* of the Corporation of Professional Engineers of Quebec comments on the revision of the 1920 Act which " . . . in its form and basic tenure have remained unchanged and it has become generally agreed among membership that, as a charter of professional rights and responsibilities, is now somewhat obsolete and requires a thorough rewriting."

The Legislation Committee of the Quebec Corporation is now preparing a complete revision of the Act and members of the Corporation have been requested to submit views and suggestions. It is hoped to be able to submit the redraft of the Act to the Quebec Legislature during the current session.

Members of the Institution in Quebec are hoping that the opportunity to re-draft the Act will permit recognition of the independent status of the communications and electronics engineer.

International Congress on Microwave Tubes

The 4th International Congress on Microwave Tubes organized by the Nederlands Radiogenootschap (NRG) and sponsored by U.R.S.I., will be held at the Technological University of Delft, Netherlands, on 3rd-7th September 1962. It will be a continuation of the three previous conferences, held in Paris 1956, London 1948 and Munich 1960.

Further information may be obtained from: The Congress Office, P.O. Box 62, Eindhoven.

Group Provident Scheme

The British United Provident Association (B.U.P.A.) is a service which operates in the United Kingdom for subscribers who wish to cover themselves for private hospital treatment instead of treatment under the National Health Service. The Institution has established a Group Scheme for members with the co-operation of the B.U.P.A. and details of annual subscription rates, medical benefit, etc., may be obtained from the Group Secretary, B.U.P.A., 9 Bedford Square, London, W.C.1. Participation in the Scheme is also open to members of European nationality who are resident in Commonwealth countries and those interested are also invited to write to the Group Secretary.

Back Copies of the *Journal*

A very prompt and large response was made by members in returning back copies of *Journals* asked for in the April issue. A large enough number was received to fulfil existing orders and provide a stock for the future. Will members therefore please note that *no further copies of these issues are required.*

621.394.346: 681.14

Considerations in the Choice of the Optimum Data Transmission Systems for use over Telephone Circuits

By

A. P. CLARK, M.A.†

Presented at the Symposium on "Data Transmission" in London on 3rd January 1962.

Summary: The transmission properties of telephone circuits are first considered and different modulation methods are then compared on the basis of their relative performance under these conditions. Phase modulation is found to be superior to other methods and the best arrangement is where the signal carrier is synchronized to the modulating waveform at the transmitter and both sidebands are transmitted. For general applications over all types of telephone circuits differentially-coherent detection has important advantages over the more conventional coherent detection. Various methods of transmitting the timing information are considered, leading to the choice of a synchronous system in which the timing waveform in the receiver is extracted from the detected signal waveform or its equivalent. 94 references.

Table of Contents

List of Symbols and Definitions

a.m. = amplitude modulation
f.m. = frequency modulation
p.m. = phase modulation

An "element" is a unit signal pulse.

A "bit" is equal to the unit of information which is the choice between two equally probable alternatives. Thus a binary element which may be either a "1" or a "0", contains one bit of information and is therefore itself often referred to as a bit.

A "baud" is the unit of modulation rate. In synchronous systems or in applications where each signal element always has the same duration, the signalling speed in bauds is equal to the number of signal elements per second.

† British Telecommunications Research Ltd., Taplow Court, Taplow, near Maidenhead, Berkshire.

1. Introduction

This discussion on the problems of data transmission over telephone circuits is an attempt to summarize as far as is possible some of the more important conclusions which have been reached on this subject, as the result of some five years work in designing and testing data transmission systems for use over telephone circuits. The bibliography at the end of this paper lists some of the more important published papers and other references whose results and opinions have also been considered in the preparation of this paper.

In order to contain the scope of this discussion within reasonable limits, we shall only be concerned here with the problems more directly related to the transmission of information over telephone circuits, and such topics as coding, system design, peripheral equipment and storage media, will not be considered. The aim of this paper is thus to describe very briefly the transmission characteristics of the different types of telephone circuits in Great Britain, as far as these are known, to examine the different known transmission systems, from the point of view of their relative performance as practical systems under these conditions and to conclude, on the basis of this comparison, which transmission systems would be the most suitable for the different practical applications.

The particular type of data transmission which is considered here is that where the information is coded in digital form. The most common arrangement of this uses binary coded signals, where each signal element may have one of two different forms or values.

2. Transmission Properties of Telephone Circuits

2.1. *The Contribution of the Individual Links to the Transmission Properties of Telephone Circuits*

A telephone circuit connecting one subscriber to another will normally be made up of two or more separate links connected in tandem. Each of these links will usually, in Great Britain, be one of three distinct types as follows: an unloaded audio link, a loaded audio link or a carrier link. Microwave and h.f. radio links are also sometimes used.

A typical longer and more complex telephone circuit could be made up of the following arrangement of individual links: unloaded audio, loaded audio, carrier, carrier, loaded audio, unloaded audio. For this reason the types of distortion to be expected over a telephone circuit will usually be a combination of the distortions introduced by the three different types of individual links.

Unloaded audio links are generally very short and have a good frequency response. Loaded audio links may be very much longer. They have the characteristic property of restricting more severely the high frequency response of the circuit and of introducing more delay distortion at the high frequency end of the band. Carrier links are in general the longest of the three types. They have the characteristic property of restricting more severely the low frequency response of the circuit and of introducing more delay distortion at the low frequency end of the band. They also often introduce a small and variable shift in the frequency location of the transmitted signal spectrum.

2.2. *Comparison of the Transmission Properties of Private and Switched Lines*

Telephone circuits themselves may be divided into two distinct groups: private and switched lines. A private line is one which is rented permanently or on a part-time basis by the subscriber. It is not connected through any of the automatic switches in the exchange and is also disconnected from the Post Office battery supplies which are used for d.c. signalling and various other purposes. The private line is also checked for its overall attenuation-frequency characteristic to ensure a reasonable overall frequency response. A switched line, that is a line on the public network, is the circuit obtained when using an ordinary telephone to set up the call, either by dialling a number or through the local exchange operator. The line is connected through a number of switches. It is connected through transmission bridges, which increase the attenuation in band and further reduce the low frequency response of the line. The line is also connected to the Post Office battery supplies, and the common impedance coupling through these to the other equipment and telephone circuits fed from them may sometimes cause considerable noise pick-up, probably more serious than that direct from the switches themselves. The noise level over switched lines is therefore in general very much higher than that over private lines. Furthermore a switched line is made up of a number of separate links, each chosen at random from sometimes quite a large number. It is therefore obviously not possible to check the frequency characteristics of complete circuits, since far too many possible combinations of the individual links would be involved, but instead only the frequency characteristics of the individual links are checked. One of the results of this is that it is possible over any such telephone circuit to obtain an unfortunate combination of the individual links, resulting in serious mismatching between these links. This will give rise to reflections in the transmitted signal, whose effect can be that of severe attenuation and delay distortions, thus seriously restricting the useful frequency band. On account of the reflections caused by mismatching, the attenuation and delay distortions experienced over switched lines may be appreciably more severe than those over private lines.

The use of mixed 2-wire and 4-wire working over private or switched lines can also introduce additional attenuation and delay distortions, due to the reflections introduced when the hybrid transformers used to change from 2-wire to 4-wire working, or vice versa, are not accurately balanced. This is one of the reasons why the high quality private lines using 4-wire working throughout have in general better frequency characteristics than private lines using 2-wire and 4-wire working.

A further disadvantage of switched lines from the point of view of data transmission is that in order to prevent false operation of the in-band voice-frequency signalling equipment connected to many of these lines in Great Britain, the frequency band from around 450 to 900 c/s cannot be used for data transmission. In some countries, such as for instance the U.S.A., no in-band v.f. signalling is used in the band below 900 c/s and the whole of this frequency band can therefore be used for data transmission, giving an appreciably wider useful bandwidth. It is of course for this reason that considerably higher signalling speeds can in general be achieved over the public telephone network in the U.S.A. than can be achieved in Great Britain.

2.3. *Attenuation and Group Delay Characteristics of Telephone Circuits*

Figure 1 shows the attenuation-frequency characteristic of an ideal voice frequency channel. Frequencies below 300 c/s and above 3000 c/s are not required for the intelligible reception of speech and the attenuation of the channel may therefore increase rapidly at frequencies below 300 c/s and above 3000 c/s.

Figure 2 shows the typical attenuation-frequency characteristic of a private line, containing both audio and carrier links. The whole of the frequency band from 300 to 3000 c/s is here available for transmission but there is a gradually increasing attenuation at frequencies above about 1000 c/s.

Figure 3 shows the typical attenuation-frequency characteristic of a switched line involving long audio and carrier links. Since in Great Britain the frequency band from around 450 to 900 c/s may not be used for data transmission, the available frequency band is limited to 900 to 2100 c/s, together with a narrow band at the low frequency end near 300 c/s. The band from 900 to 2100 c/s has an attenuation distortion of some 22 dB and this obviously places some rather severe restrictions on the maximum rate of transmission which may reliably be obtained over these lines.

Figure 4 shows a typical group-delay-frequency characteristic for a private line containing both audio and carrier links. As would be expected the group delay increases towards the lower and upper ends of the frequency band, there being a delay distortion of

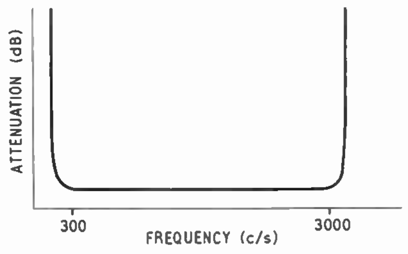


Fig. 1. Idealized attenuation—frequency characteristic for a good private line.

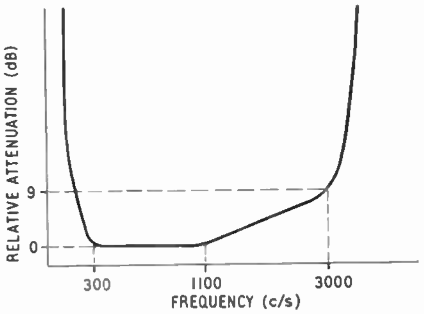


Fig. 2. Typical attenuation—frequency characteristic for a private line containing both audio and carrier links.

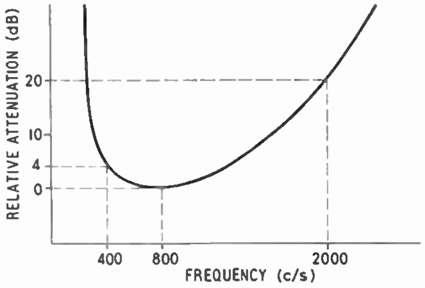


Fig. 3. Typical attenuation—frequency characteristic for a switched line involving long audio and carrier links.

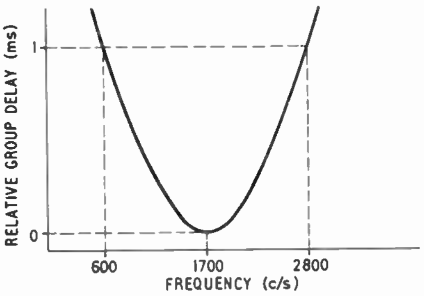


Fig. 4. Typical group-delay-frequency characteristic for a private line containing both audio and carrier links.

about one millisecond in the frequency band 600 to 2800 c/s. The better private lines tend to have a smaller increase in group delay, particularly at the lower frequencies, say only ½ millisecond at 600 c/s.

Figure 5 shows the typical group-delay-frequency characteristic for a switched line involving long audio and carrier links. This characteristic shows a considerable increase in the group delay at higher frequencies, this being contributed of course by the long audio links. The characteristic shows a delay distortion of two milliseconds in the frequency band 300 to 2400 c/s, and the presence of this distortion, particularly at the higher frequencies, places a further restriction on the maximum rate of transmission which may be obtained over the public network.

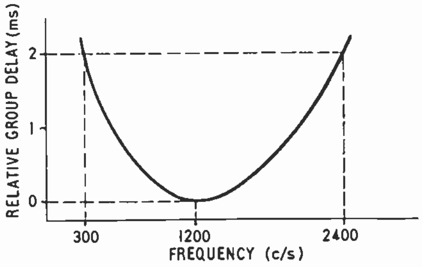


Fig. 5. Typical group-delay-frequency characteristic for a switched line involving long audio and carrier links.

The five characteristics just considered have been given as typical examples of the characteristics of telephone circuits, and although it is unlikely that frequency characteristics very much worse than those of Figs. 3 and 5 will be experienced in Great Britain, there is a very wide spread in the different frequency characteristics obtained, between those showing very little distortion and those similar to Figs. 3 and 5. The frequency characteristics also illustrate the considerable increase in both attenuation and delay distortions which may be experienced over switched lines as compared with private lines.

2.4. *Characteristics of the Noise on Telephone Circuits*

The noise obtained on telephone circuits may be classified into two distinct groups: additive noise in which a waveform is added to the transmitted signal, and multiplicative noise in which the transmitted signal is modulated by the interfering waveform. The multiplicative noise itself may be further subdivided into amplitude and frequency modulation effects.

When the interfering waveform reaches a sufficient level, its effect is to cause the receiver to interpret incorrectly the received signal and so of course to introduce errors at the receiver output. Additive noise becomes less effective in producing errors as the signal level is increased, whereas multiplicative noise has the same effect in producing errors regardless of the signal level.

The types of additive noise experienced over telephone circuits are impulsive noise, speech and signalling tone crosstalk, and white noise. Impulsive noise is the predominant type of noise over switched lines where its effects will often swamp those of the other types of noise. It is also sometimes important over private lines. Speech and signalling tone crosstalk are in general not important in causing errors except at unusually low signal levels or under definite line fault conditions. White noise will also only produce errors at very low signal levels and is not normally a significant source of errors.

The amplitude modulation effects experienced over telephone circuits are modulation noise, transient interruptions and sudden signal level changes.

Over many private lines containing carrier links and particularly at the higher signal levels, modulation noise will probably cause a large number of the errors. This appears as amplitude modulation of the signal by band-limited white noise and it normally occurs in short bursts, each lasting up to one or two seconds and causing a scattered group of errors over its duration.

Transient interruptions appear as complete breaks in transmission lasting usually from 1 to 100 milliseconds. Over some of the longer and more complex private line telephone circuits these transient interruptions could be responsible for a considerable number of the errors obtained, although over the shorter telephone circuits in Great Britain they are not likely to be important.

Sudden signal level changes, usually of the order of 1 or 2 dB but sometimes even exceeding 5 or 6 dB, may occur several times a day. Again this effect will tend to become more frequent and serious over the longer and more complex telephone circuits, and to be less important over the shorter circuits.

The frequency modulation effects experienced over telephone circuits are gradual frequency drift and sudden frequency fluctuations. These effects occur of course only over telephone circuits containing carrier links, and they correspond to a shift in the frequency location of the received signal spectrum.

In Great Britain the gradual frequency drift may typically vary from 0 to ± 5 c/s. In one or two other countries this frequency drift may be more severe and may typically reach ± 20 c/s.

The sudden frequency fluctuations have been observed in Great Britain and in one or two other countries. They may occur quite regularly and each last for a short period probably not much exceeding a millisecond or two. During this period there is quite a large frequency excursion which may be of the order of ± 100 c/s or more and which causes a large and rapid phase shift in the signal carrier.

Thus the noise over telephone circuits may be classified into three distinct groups: additive noise, amplitude modulation effects and frequency modulation effects, whose various component parts are of the type outlined above. Switched lines have the same amplitude and frequency modulation effects and additive noise as private lines, but they have in addition a high level of impulsive noise which, as previously mentioned, may often mask the other types of noise present.

Experimental and theoretical considerations have shown that although the tolerance of a data transmission system to white noise is not necessarily an accurate measure of its actual tolerance to the additive noise over telephone circuits, the relative tolerance to white noise of different data transmission systems is nevertheless a good measure of their relative tolerance to this additive noise. Since white noise lends itself well to theoretical calculations and is also easily produced in the laboratory over the required frequency range, the relative tolerance of different data transmission systems to white noise is a very useful measure of their relative tolerance to the additive noise over telephone circuits.

Whereas over switched lines the majority of the noise is additive, over private lines amplitude and frequency modulation effects tend to predominate, and these must therefore also be considered. Thus the relative performance of different data transmission systems over telephone circuits may be assessed on the basis of their tolerance to white noise and to both amplitude and frequency modulation effects, due weight being given to the relative frequency and severity of the different types of noise represented by these.

3. Modulation Methods

3.1. *The Need for Modulation*

The information which it is required to transmit over the telephone circuits is most often initially contained in a binary coded square wave signal in which one voltage level represents "1" and another voltage level represents "0". The main function of the data transmission equipment is at the transmitter to convert this square wave signal into the form most suitable for transmission over telephone circuits and at the receiver to convert the latter waveform back again into the original square wave signal.

The information carrying square wave signal cannot itself be transmitted satisfactorily over the typical telephone circuit for two good reasons. Firstly, whatever the signal element rate, a large proportion of the transmitted energy will be lost, and secondly, such energy as does reach the other end of the telephone circuit will normally be so severely distorted as to make satisfactory detection impossible. For this reason the information carrying square wave signal must always be used to modulate a sine wave carrier in such a way that the bulk of the energy is shifted from the lower frequencies to the best part of the available frequency band.

In cases where a private line telephone circuit is less than about 15 miles long and is not connected through a repeater or any telephone carrier equipment, it may be possible to obtain a circuit giving a d.c. connection between the two ends and passing all the low frequencies with little or no distortion. Under these conditions a suitably rounded form of the information carrying square wave signal could be transmitted directly over the circuit. Such an arrangement should give satisfactory operation and almost certainly the highest obtainable signalling speed for binary coded signals over the particular type of circuit.

In order to simplify the following discussion it will be assumed throughout that only binary coding is being used. Many of the same basic principles will of course also apply where multi-level coding is used, but the detail involved becomes more complex.

3.2. *Multi-frequency Systems*

Over any transmission path having a given bandwidth, the transmitted signal may be either a single modulated carrier which carries all the transmitted information serially and occupies all the available frequency band, or it may be made up of several different modulated carriers each of which carries part of the transmitted information and occupies part of the available frequency band. In general the former approach results in a considerable saving of equipment when compared with the latter, largely because of the inevitable duplication of equipment in any multi-frequency system, usually at the receiving end where the different frequency channels must be separated from each other. The one application where there is sometimes less equipment involved in multi-frequency systems, is a data collection system which may use one or two receivers fed from any one or two of say a hundred different transmitters. The complexity of the receivers is here obviously unimportant compared with that of the transmitters, and under these conditions a multi-frequency system need not be under any disadvantage on grounds of economy. Again under particular conditions where special types of interference and signal distortion are found, such as over short wave radio links, and where cost is relatively unimportant, multi-frequency systems may in some cases be considered more suitable. In general, however, for applications over telephone circuits, a data transmission system using a single modulated carrier is nearly always to be preferred on grounds of both economy and performance to a multi-frequency system.

3.3. *A.M., F.M. and P.M. Systems*

A sine wave carrier may be expressed by the equation

$$y = A \sin (2\pi ft + \phi)$$

This waveform contains three different parameters which are variable quantities and which may be used to carry information. These are the amplitude A, the frequency f and the phase ϕ. In any simple process of modulation, one of these parameters is made to vary with the modulating waveform so that the information present in the modulating waveform is transferred to that parameter of the sine wave carrier. This process is illustrated in Fig. 6. The modulating waveform here is the binary coded square wave signal which contains two adjacent elements, the first of which represents a "0" and the second a "1".

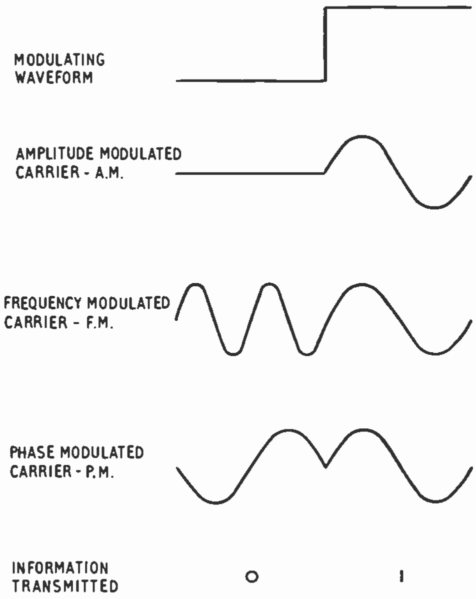

MODULATING WAVEFORM
AMPLITUDE MODULATED CARRIER - A.M.
FREQUENCY MODULATED CARRIER - F.M.
PHASE MODULATED CARRIER - P.M.
INFORMATION TRANSMITTED
0
1

Fig. 6. Basic types of modulation for a binary digital system.

Where 100% amplitude modulation is used, the first signal element, which corresponds to a "0", is represented by the absence of the signal carrier, and the second signal element, which corresponds to a "1", is represented by the presence of the carrier.

Where frequency modulation is used, a "0" is represented by one frequency of the sine wave carrier and a "1" is represented by a different frequency.

Where phase modulation is used, the phase of the sine wave carrier in one signal element is shifted by 180 deg relative to that in the other. Where each signal element contains an exact multiple of one cycle of the signal carrier, the sine wave carrier in one signal element is also inverted with respect to that in the other.

In practical phase modulated transmission systems using a binary coded signal, a "1" cannot simply be represented by a given phase and a "0" by the inverse phase, because in transmission the carrier phase may be shifted by a large and often variable amount, so that an element transmitted as a "0" may well arrive at the other end as a "1" and vice versa. To overcome this difficulty a "1" is normally transmitted as a 180 deg phase shift in the carrier between two adjacent signal elements and a "0" as no phase shift between these elements. Alternatively the phase of the received signal carrier corresponding to a "0" or a "1" may be determined at the beginning of each transmission by sending here a continuous series of "0's" or "1's" or some other convenient code, and then the phase of the carrier corresponding to a "0" or a "1" at any later point during transmission may be remembered by suitable means, so long as the phase variations of the carrier, resulting from shifts in the frequency location of the received signal spectrum, are not rapid enough to be confused with a real transmitted phase inversion between adjacent signal elements. Again there are other arrangements, involving for instance either the additional transmission of a reference carrier, the use of a phase shift of less than 180 deg, or restrictions in the transmitted signal code, whereby the receiver is enabled to identify the phase of the received signal carrier corresponding to a "0" or a "1" at any point during transmission.

3.4. *Two Errors to be Avoided*

One mistake which is sometimes made in comparing different modulation methods, is to assume that the relative tolerance to noise for different types of modulation, in the case of data transmission systems using telephone circuits, is similar to that for other systems such as broadcast radio transmission. For instance the normal f.m. broadcast radio transmission system achieves a considerably higher tolerance to low-level noise than does the equivalent a.m. system. However it does this essentially by using a very much wider frequency band. Over telephone circuits, where the available bandwidth is severely limited, data transmission systems must normally be designed to operate at the maximum possible speed over the given bandwidth, and at these speeds of working f.m. systems can therefore achieve no advantage at all in tolerance to noise over a.m. systems, by virtue of a wider frequency band.

Another mistake which is perhaps more frequently made, is to neglect the fact that, whereas in the typical radio signal or in 50 or 75-baud line telegraphy, there are many cycles of the signal carrier for each signal element, in data transmission over telephone circuits there are normally only one or two cycles. Some calculations on the relative tolerance to noise of the different modulation methods assume the fact that the carrier frequency is considerably higher than the highest modulating frequency, in other words that there are many cycles of the signal carrier per signal element. Where the results of such calculations are appreciably affected by this assumption, the results obviously do not apply to the conditions existing for data transmission over telephone circuits.

The concept of an "ideal detector", for instance, which assumes a detector whose output voltage or current is linearly proportional to the instantaneous amplitude, frequency or phase of the received signal carrier, although reasonably closely approached in the case of practical detector circuits in radio transmission or 50-baud line telegraphy, is by no means so closely approximated in the case of data transmission over telephone circuits. In the latter case, the discrepancy between the performance of the ideal detector and that of a practical detector circuit may vary widely from one modulation method to another and from one particular receiver design to another, there being in some cases a serious degradation in the performance of the practical detector circuit when compared with that of the ideal detector.

In general, for data transmission systems designed to work over telephone circuits, the signals transmitted and the methods of detection which must be used are such that the performance of the receiver is much more accurately described in terms of a device which must determine for each received signal element whether its shape corresponds more closely to one or other of the two correct binary forms, rather than a device which responds solely to the instantaneous amplitude, frequency or phase of the received signal carrier.

3.5. *Cross-Correlation Coefficient*

The theoretical approach which seems to correspond most closely to the action of the most common detector circuits in f.m. and p.m. systems is that based on the cross-correlation coefficient of the two binary elements in the signal waveform used.

If the time duration of a signal element is T seconds and if the shape of one of the two binary elements is described by the real function $f_1(t)$ in the time interval 0 to T seconds and the shape of the other is described by the real function $f_2(t)$ in the same time interval, then the cross-correlation coefficient of these two elements is defined by the expression

$$\frac{1}{E}\int_0^T f_1(t).f_2(t)\,dt$$

where E is the square root of the product of the energies in the two binary elements. In the case of f.m. and p.m. systems, where the energy in either element is normally the same, E is equal to the energy of either signal element.

The cross-correlation coefficient is simply the integral of the product of the two binary elements, taken over the corresponding points in these two elements, this integral being normalized by multiplying by the factor $1/E$, so that it has a maximum positive or negative value of unity. The cross-correlation coefficient gives a measure of the difference between the shapes of the two binary elements, by means of which difference the detector is able to discriminate between one and the other. When the cross-correlation coefficient is equal to +1, the shapes of the two binary elements are the same, and the detector cannot therefore use the shapes to discriminate the one from the other. When the cross-correlation coefficient is equal to −1, there is the maximum difference between the shapes of these elements, the one being the negative or inverse of the other. The nearer the cross-correlation coefficient is to −1, the greater is the useful difference between the two binary elements and therefore the greater the mutilation or distortion which can be tolerated by either of the two elements before it can no longer be distinguished from the other. Thus the nearer the cross-correlation coefficient is to −1, the more noise and signal distortion can be tolerated by the system.

When there is no correlation or coherence between the two binary elements, in other words if the relationship between the two elements is such that it cannot be said that their shapes are more nearly the same or more nearly different, then the cross-correlation coefficient is equal to zero.

In the case of an f.m. signal, the cross-correlation coefficient is greater than or equal to zero and can never have a negative value. The reason for this is that when two different carrier frequencies are used for the two binary elements, any phase coherence or correlation there may be between the two binary elements at the transmitter, will in general be lost by the time the signal reaches the receiver. After transmission over telephone carrier circuits, the phase of one carrier frequency relative to that of the other may also be continuously varying with time. Under these conditions, an examination of the cross-correlation coefficient shows that its true value is always greater than or equal to zero.

For a rectangular modulating waveform as shown in Fig. 6, when the frequency shift (in c/s) between the two different carrier frequencies is an exact multiple of the signal element rate or bit rate, the cross-correlation coefficient is exactly equal to zero, and this is therefore the optimum condition for the f.m. system. The narrowest transmitted frequency band, for the optimum value of the cross-correlation coefficient, is obtained when the frequency shift (in c/s) is exactly equal to the transmitted bit rate. Quite often in order to reduce further the frequency band of the signal, so as to obtain the maximum possible rate of transmission, a somewhat smaller frequency shift is used. Under these conditions the cross-correlation coefficient has a positive value, and there is a corresponding reduction in the tolerance of the system to noise and distortion.

As shown in Fig. 6, the relative phase of the signal carrier in the two binary elements of a p.m. system, is normally 180 deg. Under these conditions and provided that the signal carrier is synchronized to the modulating waveform so that each signal element contains an exact multiple of one half cycle of the signal carrier with at least one complete cycle, one of the two binary elements is always the inverse of the other and the cross-correlation coefficient is therefore equal to -1. Such a system has the greatest possible tolerance to noise and signal distortion, by virtue of the particular shape of the signal elements used. The fact that the phase of the signal carrier may be shifted by a large and variable amount in transmission over telephone circuits is of no importance here, since the phase shift experienced by the signal carrier in either of the binary elements is always essentially the same. Thus whatever the phase of the signal carrier, its relative phase in the two binary elements will be 180 deg, thereby always maintaining the value of the cross-correlation coefficient at -1.

In order to simplify the above explanation, the phase of the signal carrier is here considered relative to the signal elements themselves. Using this measure of the carrier phase, two elements having the same carrier phase will always be identical and two elements having a relative carrier phase of 180 deg will always be inverted with respect to each other, provided only that the carrier is synchronized to the modulating waveform at the transmitter, in the manner specified above. However, when there is an odd multiple of half cycles of the carrier per signal element, two adjacent elements which are identical in fact involve a 180 deg phase shift in the carrier at the junction between these elements, and two adjacent elements which are inverted with respect to each other in fact involve no phase change in the carrier here. On the other hand, when there is an exact multiple of one cycle of the carrier per signal element, as for instance in Fig. 6, two identical elements always in fact involve no phase change in the carrier and two elements which are inverted with respect to each other always in fact involve a 180 deg phase shift in the carrier.

3.6. *Relative Tolerance to White Noise of A.M., F.M. and P.M. Systems*

On account of the difference in the cross-correlation coefficients and comparing the optimum arrangements, a p.m. system has a 3 dB advantage in tolerance to white noise over an f.m. system. In addition, because in a p.m. system synchronous or coherent detection is used, whereas in an f.m. system envelope or non-coherent detection is normally used, the p.m. system gains another 1 dB advantage. On account of the two factors just mentioned, therefore, an ideal p.m. system gains a 4 dB advantage in tolerance to white noise over an ideal f.m. system.

These results apply at the lower error rates of around 1 bit in 10^5. As the error rate increases so the advantage in tolerance to white noise of the ideal p.m. system increases, reaching 6 dB at an error rate of 1 bit in 10. In the following discussion we shall however only consider the more practical situation where the error rate is around 1 bit in 10^5.

The advantage in tolerance to white noise of one system over another is here expressed as the difference in decibels between the normalized signal/noise power ratios for the two systems when these are giving the same error rate. The normalized signal/noise power ratio is the signal energy per bit divided by the noise power density, or it may alternatively be expressed as the ratio of signal power to the signalling speed in bits per second, divided by the ratio of noise power to the receiver bandwidth.

Over telephone circuits a limit is set on the peak signal power which may be transmitted to line. On account of this the mean power level for both f.m. and p.m. signals will normally be 3 dB higher than that for an a.m. signal, assuming that in the a.m. signal each of the two binary elements has the same probability of occurrence. Under these conditions and at an error rate of around 1 bit in 10^5, a p.m. system has a 7 dB advantage over an a.m. system in tolerance to white noise. This is a well known fact which has been verified independently by both theoretical and practical means. Of these 7 dB, 6 dB are contributed by the difference between the two waveforms used and 1 dB by the fact that an a.m. system normally uses envelope detection in place of synchronous detection as used in the p.m. system.

Thus taking all the above mentioned factors into consideration, a p.m. system has a 4 dB advantage over an f.m. system in tolerance to white noise and an f.m. system has a 3 dB advantage over an a.m. system.

3.7. *Relative Tolerance to Multiplicative Noise of A.M., F.M. and P.M. Systems*

Modulation noise, which is one of the more important components of the multiplicative noise over telephone circuits, has probably much the same relative effect on the different modulation methods as white noise, and therefore in tolerance to this type of noise a p.m. system will give the best performance and an a.m. system the worst with an f.m. system about halfway between.

Transient interruptions will of course affect all modulation methods to a similar degree.

Sudden signal level changes will have by far the most serious effect on a.m. systems and may in some cases seriously impair their performance. Neither f.m. nor p.m. systems when correctly designed should however be affected by these.

Frequency modulation effects, which form the remaining two components of the multiplicative noise, will have the most serious effect on p.m. systems and the least effect on a.m. systems. The tolerance of an a.m. system to frequency modulation effects is appreciably better than that of an f.m. system and the tolerance of the latter is in turn considerably better than that of a p.m. system. Both a.m. and f.m. systems have in general a very adequate tolerance to the frequency modulation effects experienced over telephone circuits, but p.m. systems of the type so far considered, using coherent detection, may be seriously affected by the more severe frequency modulation effects, such as those which may occur over some of the older telephone carrier circuits.

The low tolerance of p.m. systems to frequency modulation effects is the one disadvantage of such systems, which would otherwise be clearly the most suitable when judged on the various other factors so far considered. Of course many of the shorter telephone circuits use only audio links, over which no change in the frequency location of the received signal spectrum can occur. Also over many carrier links the frequency shift in the received signal is in practice limited to just one or two c/s, and in this case it is a relatively simple matter to design a p.m. receiver to tolerate these frequency shifts with no degradation in performance. Over these two types of telephone circuits the lack of tolerance of a p.m. system to frequency modulation effects is of no importance and a p.m. system is undoubtedly the most suitable when judged on tolerance to noise and distortion.

However for more general applications particularly over some of the older telephone carrier circuits, the frequency shifts experienced from time to time will be more than those which can be tolerated by a p.m. system of the type so far considered, and under these conditions such a system may therefore give a poorer overall performance than say the equivalent f.m. system.

3.8. *P.M. Systems using Coherent and Differentially-Coherent Detection*

The previous discussion on the relative tolerance to additive and multiplicative noise of the different modulation methods, has assumed throughout that in the case of the p.m. system, coherent detection is used in the receiver, this being perhaps the more conventional design. In coherent detection the received signal is detected by comparing it in a phase-sensitive detector with a reference sine wave. This sine wave must itself be derived from the received signal and is arranged to be in phase with one or other of the two different binary elements in this signal. In order that the reference sine wave may be undistorted and relatively free from noise, it must, whether obtained directly from the received signal or indirectly via a phase locked oscillator, be limited to a narrow frequency band and therefore only capable of a relatively low rate of change of phase. As a result of this, when the received signal spectrum undergoes a significant frequency excursion, causing a rapid phase variation in the signal carrier, the reference sine wave is not able to maintain the correct phase relationship with the received signal carrier, and inefficient or faulty detection will always result. The typical p.m. receiver using coherent detection will tolerate an excursion in the frequency location of the received signal spectrum of up to around ± 10 or 20 c/s. Such a p.m. system will therefore inevitably be at a serious disadvantage under the more severe conditions of frequency modulation experienced over telephone circuits. It is for this reason of course that synchronous or coherent detection is not normally used in a.m. or f.m. systems, since the 1 dB advantage gained in tolerance to white noise is more than offset by the very serious reduction in tolerance to frequency modulation effects.

There is however an alternative approach to this method of detection of phase modulated signals and this is known as differentially-coherent detection. This method effectively overcomes the one serious disadvantage of coherent detection, by increasing the tolerance to frequency modulation effects by a factor of at least some five or ten times. In differentially-coherent detection, the reference waveform with which the received signal is compared in the phase sensitive detector, is the received signal itself after this has been delayed in most cases by the exact duration of one signal element. Each signal element is thus directly compared with that immediately preceding it. In such a system a "1" is also normally transmitted as a phase shift of 180 deg in the carrier between adjacent signal elements and a "0" as no phase shift between these elements. If at the end of this comparison of any pair of signal elements it is decided that the two were on the average more nearly of the same phase, a "0" is indicated, and if it is decided that they were on the average more nearly 180 deg out of phase, a "1" is indicated. The received signal is thus not only detected but also reconverted to the original code in which a "1" is represented by one of the two binary elements and a "0" by the other. Because each signal element is directly compared with that immediately preceding it, frequency modulation of the transmitted signal can only cause faulty detection when the mean phase of the signal carrier in one of the two signal elements is shifted ideally by at least 90 deg relative to that in the adjacent element. When each signal element contains only one or two cycles of the signal carrier, this obviously corresponds to a considerable frequency excursion and certainly a very much larger frequency excursion than could be tolerated by the equivalent coherent detector. A differentially-coherent detector will in fact ideally only produce an error in the output signal when the received waveform itself is in error, and under these conditions no detection system could do better.

Another great advantage of differentially-coherent detection is that correct detection is obtained after receiving only one signal element following a break in transmission, whereas with coherent detection there is always an appreciable delay required, possibly even exceeding 100 milliseconds, before the reference sine wave has both the right amplitude and phase for correct detection. In those practical applications where it is required that a signal is only fed to line during the transmission of an actual message, no-signal is used as a stand-by condition between messages, and the need for an appreciable delay before useful information may be transmitted, following the no-signal condition, may in some cases be a serious disadvantage. Coherent detection used under these conditions is under a similar disadvantage to an a.m. system using automatic gain control in the receiver.

A further practical advantage of differentially-coherent detection is that because it does not involve the precise adjustment of tuned circuits in order to obtain the correct phase of the reference sine wave, it is inevitably very much less liable to inferior performance due to inaccurate initial adjustments or drift with temperature and time. The only frequency sensitive element used in differentially-coherent detection is a delay network, whose stability may easily be made more than adequate even under the worst practical conditions. The practical circuitry involved in differentially-coherent detection is therefore capable of considerably greater reliability than that in coherent detection and it is also, if anything, somewhat simpler.

The disadvantage of differentially-coherent detection when compared with coherent detection, stems from the fact that, whereas in coherent detection the received signal is compared with a relatively pure sine wave which has been filtered to remove both noise and distortion, in differentially-coherent detection each signal element is compared with an adjacent element, where each of these are in the presence of both noise and distortion. Intuitively therefore one would expect the tolerance to white noise of a p.m. system using differentially-coherent detection to be degraded by 3 dB relative to a p.m. system using coherent detection. Theoretical analysis and practical measurements have however both shown that for transmission systems using binary coded signals the reduction in tolerance to white noise is only about 1 dB for typical error rates of around 1 bit in 10^5. For systems using quaternary and other multi-level codes, on the other hand, the degradation is in fact more nearly 3 dB as would have been expected.

In a practical coherent detector, as the effective filtering applied to the reference waveform is made finer, to reduce the level of the noise present, so the tolerance to frequency modulation effects on the received signal is reduced, and in practice a compromise must be reached between the two requirements. One result of this is that a practical coherent detector cannot approach so closely to the ideal coherent detector, as a practical differentially-coherent detector can approach to its ideal. Thus the difference in tolerance to white noise between a practical coherent detector and a practical differentially-coherent detector, for a given error rate of around 1 bit in 10^5 and under equivalent conditions, is somewhat less than 1 dB, probably more nearly ½ dB.

In order that a cross-correlation coefficient equal to -1 should be obtained for the transmitted waveform, with consequently the optimum tolerance of the differentially-coherent detector to both noise and distortion, the signal carrier must be synchronized to the modulating waveform at the transmitter output so that each signal element contains an exact multiple of one half cycle of the carrier with at least one complete cycle. Where for any reason this synchronization cannot be achieved, the value of delay used in the differentially-coherent detector should equal the duration of the largest number of half cycles of the signal carrier whose total duration is less than that of one signal element. Under these conditions the tolerance to noise and distortion will always be less than that of the ideal arrangement, because now only a portion of each signal element is effectively used in the detection process. Where coherent detection is used of course, there is no reduction in tolerance to noise and distortion on this account, since with coherent detection the whole of each element is always effectively used in the detection process. When the signal carrier is not synchronized to the modulating waveform at the transmitter, coherent detection therefore achieves a greater advantage in tolerance to white noise over differentially-coherent detection, although the improved tolerance would normally be small, not often exceeding 1 dB.

The tolerance to signal distortion of a differentially-coherent detector is also not as good as that of a coherent detector, although in practice where double sideband transmission is used the difference is not serious enough to be a significant factor in affecting the choice between the two methods of detection. However where vestigial sideband transmission is used, there may sometimes be a significant increase in signal distortion at the detector output when differentially-coherent detection is used in place of coherent detection. This is because coherent detection tends to eliminate the distortion component known as the quadrature component in the received vestigial sideband signal, whereas differentially-coherent detection will be to some degree adversely affected by it.

This advantage of coherent detection is however more than offset by the fact that a coherent detector can in general only extract the necessary information from the received signal in order to reconstitute the reference carrier needed for detection, from both the upper and lower sidebands in those signal patterns for which no carrier frequency is present. Since the frequency components in the upper sideband corresponding to the higher modulating frequencies are removed by the vestigial sideband filter, the reference carrier can only be derived from the lower modulating frequencies in these signal patterns. Signal patterns containing no carrier frequency and only the higher modulating frequencies, such as successive phase reversals, cannot therefore be used for regenerating the reference carrier. Thus a restriction must be placed on the transmitted code in order to limit the periods during which only the higher modulating frequencies are received in those signal patterns for which no carrier frequency is present. This is of course not only a severe limitation to the use of coherent detection with vestigial sideband p.m. signals, but it also necessarily involves either an even lower tolerance to frequency modulation effects or a lower tolerance to additive noise and distortion.

The need for restricting the transmitted signal code with vestigial sideband transmission can be avoided by adding to the transmitted signal a sine wave at the carrier frequency and suitably phased relative to the signal carrier, whereby the receiver is enabled under all conditions to obtain the reference carrier needed for coherent detection. However, such an arrangement is inevitably more complex, due to the requirement for preventing interference between the transmitted signal and the additional carrier, and it has a lower tolerance to noise. The receiver here is also normally sensitive to amplitude modulation effects such as sudden signal level changes, thus adding a further and more serious weakness to this method of coherent detection.

The great advantage of differentially-coherent detection in this application, is that it does not require to derive the signal carrier frequency from the received signal in order to achieve correct detection. The reason for this is simply that a differentially-coherent detector uses the received signal waveform itself, after this has been suitably delayed, as the reference waveform for the phase sensitive detector. Thus as long as a signal is received, the correct reference waveform will also be obtained with the correct relative time delay. Tests on a practical vestigial sideband p.m. system using differentially-coherent detection have shown conclusively that the detector in such a system will work correctly for all signal patterns and with no serious degradation in performance, not only in the complete absence of frequency modulation effects but also for appreciable shifts in the frequency location of the transmitted frequency spectrum on either side of its correct value. This applies either with or without the signal carrier synchronized to the modulating waveform at the transmitter and for all carrier frequencies greater than and not much less than the signal element rate.

However, as mentioned previously, a p.m. system using differentially-coherent detection experiences under all conditions a certain reduction in tolerance to noise when used with a vestigial sideband signal, due to the presence of the quadrature component in the received signal. Thus with p.m. systems using either coherent or differentially-coherent detection, double sideband transmission should always be used where permitted by the signalling speed required.

3.9. *Disadvantages of Automatic Gain Control*

Because of the wide range of signal attenuations experienced over telephone circuits, a means must normally be used to control the output signal level obtained from the input amplifier of any receiver, whatever modulation method is being used. Automatic gain control must always be used in an a.m. receiver and may also be used for f.m. or p.m. receivers. Alternatively in f.m. or p.m. receivers the received signal may be sliced in an amplifier-limiter, since with f.m. or p.m. signals the information is carried in the zero level crossings of the signal carrier rather than in the peaks or in the instantaneous carrier amplitude.

If the a.g.c. used in an a.m. receiver is not to degrade unduly the tolerance to the majority of the noise experienced over telephone circuits, the time constant must be kept fairly long. Under these conditions sudden signal level changes can seriously degrade the receiver performance. However, even if the time constant were made as short as possible, the sudden level changes experienced over telephone circuits would in general be too rapid for the a.g.c. to have taken effect before at least one or two errors had occurred. Furthermore, since the duration of the noise bursts, particularly over private lines, varies over so wide a range, typically from around one millisecond to several seconds, whatever the time constant of the a.g.c. amplifier used in the receiver, there will always be some noise pulses which, because of their short duration, will not affect the signal amplitude at the output of the a.g.c. amplifier. There will also always be noise bursts which will have a sufficiently long duration to reduce this signal level appropriately for the added presence of the noise. Thus if the receiver is adjusted for optimum detection under conditions of no-noise, towards the end of a long burst of noise the reduction in signal level at the output of the a.g.c. amplifier will inevitably cause less efficient detection and therefore a lower tolerance to this noise. Conversely if a receiver is adjusted for optimum detection in the presence of a predetermined noise level, then under conditions of short isolated noise pulses or at the beginning of a long noise burst the detection efficiency will not be the optimum, again resulting in a lower tolerance to this noise. Moreover, the higher the noise level relative to that of the signal, the greater the effect of the noise on the signal level at the output of the a.g.c. amplifier and therefore the more serious the degradation in tolerance to the noise. Thus, because in an a.m. receiver a.g.c. must be used and because in this system the information is carried in the amplitude of the transmitted signal carrier, an a.m. receiver always has a reduced tolerance to either or both the longer and shorter noise bursts. F.m. and p.m. receivers using a.g.c. are also similarly although less seriously affected in those cases where the detector output signal is in any way dependent on the amplitude of the signal carrier at the detector input. Where this occurs, of course, the tolerance of the receiver to sudden signal level changes is also correspondingly reduced.

Where slicing is used in place of a.g.c. for practical f.m. or p.m. receivers, even under conditions of no signal level changes, an improvement may generally be obtained in the tolerance of these systems to noise. Thus the fact that f.m. and p.m. systems can use slicing instead of a.g.c. to control the received signal level, gives these systems yet a further advantage over a.m. systems.

Another serious disadvantage of a.g.c. is that a certain minimum time interval is required for the a.g.c. circuits to settle down to their correct condition, after the transmitted signal is first received following a no-signal condition on the line. In some applications this places a severe limitation on the usefulness of the data transmission system, particularly where it is required to use "no-signal" as a standby condition.

For the reasons considered above, the use of a.g.c. in a receiver should always be avoided whenever possible, and therefore it should never be used with either f.m. or p.m. systems, except where other special requirements make its use unavoidable.

3.10. *Disadvantages of A.M. Systems*

An a.m. system, when used over telephone circuits where compandors are installed, is in general at a disadvantage when compared with f.m. or p.m. systems, essentially because the a.m. signal is not a continuous waveform as are the f.m. and p.m. signals.

Another disadvantage of a.m. systems, when compared with f.m. or p.m. systems, is that because the latter two systems use a continuous transmitted signal, the presence or absence of this signal at any time may with these systems be used to carry an additional piece of information, namely as to whether or not a fault condition causing loss of signal develops in the transmission path or at the transmitter, during the transmission of a message. Transient interruptions on the line exceeding about ten milliseconds may also conveniently be detected in this way. In a.m. systems, however, the absence of the signal carrier is one of the two binary elements transmitted. Thus the absence of the signal at the receiver input can here only be taken to imply a true loss of signal, if its duration exceeds the maximum obtainable with the code used for the transmitted signal, and this may often correspond to many signal elements.

An a.m. system therefore has the following disadvantages when compared with f.m. and p.m. systems. A lower tolerance to additive noise, a very much lower tolerance to sudden signal level changes, the necessary use of a.g.c. with its consequent disadvantages, a poorer relative performance over compandors and finally the inability to use efficiently the absence of the received signal carrier as an additional piece of information to indicate the true loss of signal.

For these various reasons f.m. or p.m. systems would normally always be preferred to a.m. systems.

3.11. *Relative Overall Tolerance to the Additive and Multiplicative Noise over Telephone Circuits of A.M., F.M. and P.M. Systems*

Considering ideal systems, a p.m. system using coherent detection has a 1 dB advantage in tolerance to white noise over a p.m. system using differentially-coherent detection, for error rates of the order of

1 bit in 10^5. The latter has a 3 dB advantage over an f.m. system and an f.m. system has a 3 dB advantage over an a.m. system.

Thus in tolerance to the additive noise over telephone circuits the order of preference for the different systems is as follows: a p.m. system using coherent detection, a p.m. system using differentially-coherent detection, an f.m. system, and an a.m. system.

In tolerance to the amplitude modulation effects over telephone circuits, an a.m. system is at a serious disadvantage when compared with the other systems.

In tolerance to the frequency modulation effects over telephone circuits, a p.m. system using coherent detection is at a serious disadvantage when compared with the other systems.

Considering therefore the overall tolerance to the additive and multiplicative noise over telephone circuits, a p.m. system using differentially-coherent detection will most probably give the best performance and is to be preferred to the other modulation methods.

4. Maximum Rate of Transmission over a Given Bandwidth for Different Binary Coded Signal Waveforms

4.1. *Ideal Maximum Transmission Rate*

Nyquist showed in 1928 that the maximum signal element rate which may be transmitted over a bandwidth B c/s, for no intersymbol interference, is $2B$ elements per second, and this is sometimes known as the Nyquist rate. Figure 7 shows the shape of the ideal and undistorted binary pulses which have a uniform energy distribution from 0 to B c/s and no energy above B c/s. These pulses when spaced relative to each other at time intervals which are multiples of $1/2B$, will cause no intersymbol interference if sampled at the central positive peaks. That is, at the positive peak of any one pulse there can be no signal voltage contributed by any of the other pulses.

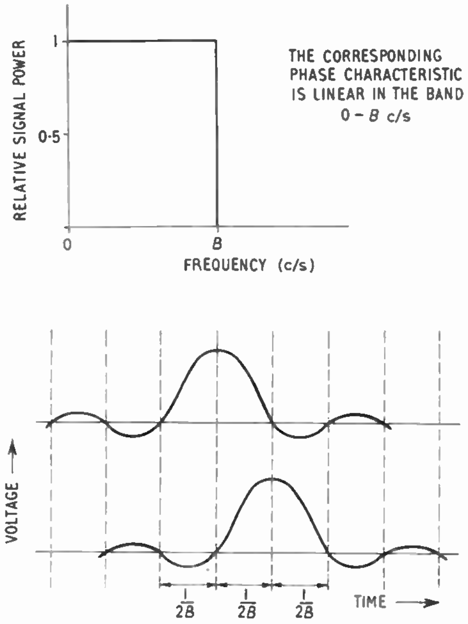


Fig. 7. Frequency spectrum and shape of the pulses with which the ideal maximum signal element rate of $2B$ elements per second may be obtained over a frequency band 0–B c/s for no intersymbol interference.

Not only is $2B$ elements per second the maximum signal element rate through a bandwidth of B c/s, for no inter-symbol interference, it is also the maximum element rate which may be used for the satisfactory transmission of all possible codes, and corresponding pulse patterns, over this bandwidth. This may be seen from the following simple example. Suppose it were required to transmit binary coded signal elements over a bandwidth 0 to B c/s at an element rate of $2C$ elements per second, where C is greater than B. Since binary coded signal elements are used, each element can have one of two different levels, and usually in this case each element will appear either as the presence or absence of a pulse. Suppose a "1" is represented by the presence of a pulse and a "0" by its absence, or alternatively that a "1" is represented by the larger of two different pulses and a "0" by the smaller. One possible and in practice frequently used code contains alternate "1's" and "0's". At an element rate of $2C$ the pulse pattern corresponding to this code must necessarily contain a frequency component of C c/s, and since C is greater than B, this cannot be transmitted over a frequency band 0 to B c/s. The same principle applies of course also with the bandwidth of B c/s located anywhere else in the frequency spectrum.

Thus the maximum signal element rate which may be transmitted over a frequency band of B c/s is $2B$ elements per second. Referring again to Fig. 7, it can be seen that in the complete absence of noise and distortion and assuming an infinitely sensitive receiver, each signal element may have one of as many different levels as required. Since the information content of each element is $\log_2 n$ bits, where n is the number of possible levels, the maximum rate of transmission of information is now $2B.\log_2 n$ bits per second, and this may be made as large as required. Thus although there is an absolute and fundamental limit to the rate of transmission of signal elements over a given bandwidth, there is no such limit to the rate of transmission of information. In practice, however, because noise and distortion are always present, the number of different levels used for the signal elements is necessarily limited, and the maximum number which may satisfactorily be used is that which gives an adequate discrimination against the effects of both noise and distortion. Since in most applications over telephone circuits the noise and distortion are the limiting factors, binary coded signal elements are used whenever possible, and multi-level coded elements are only used when absolutely necessary on account of the speed requirements. When the signal elements are binary coded, the maximum rate of transmission of information over a frequency band of B c/s, is of course $2B$ bits per second.

In order to simplify the following discussion, it will be assumed throughout that binary coded signals are being used. The various results quoted here therefore apply specifically to binary coded signals.

4.2. *D.C. Pulses*

Figure 8 shows the frequency spectra of two different d.c. pulses. A rectangular pulse of duration T seconds has the frequency spectrum shown by the dotted line. Although much of the energy is concentrated at the lower frequencies below $(1/T)$ c/s, a considerable amount of energy is distributed also over the higher frequencies and this pulse has in fact ideally an infinitely wide frequency spectrum. If the pulse is now suitably rounded so that it becomes a "sine^2" or a "raised-cosine" pulse, the energy at the higher frequencies is very much reduced and the frequency spectrum becomes as shown by the other curve. The frequency spectrum of the rectangular pulse has points of zero energy at frequencies which are multiples of $1/T$, and peaks of energy between the adjacent zeros. The frequency spectrum of the raised-cosine pulse also has some energy at frequencies above $1/T$ but at a much lower level, and this energy may often for practical purposes be neglected.

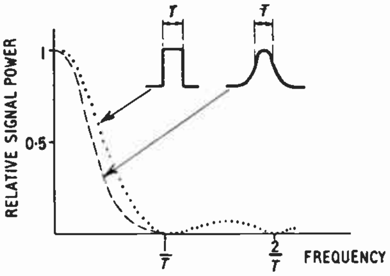


Fig. 8. Frequency spectra of two commonly used modulating waveforms for binary coded digital signals.

For a pulse or element width of T seconds the rate of transmission is $1/T$ elements per second. From Nyquist, this requires a minimum frequency band of 0 to $(1/2T)$ c/s in the transmission path, for the correct identification or detection of the separate transmitted signal elements, and thus for either of the d.c. pulses the frequency band 0 to $(1/2T)$ c/s contains the essential or useful information.

Waveforms made up of either of the pulse forms considered here may be used to modulate a suitable carrier frequency, using amplitude, frequency or phase modulation.

4.3. *A.M. Signals*

Figure 9 shows the frequency spectrum of an amplitude modulated carrier, using a binary square wave modulating waveform, with an element width of T seconds. The frequency spectrum is symmetrical about the carrier frequency f_c, the lower sideband being inverted with respect to the upper sideband and the upper sideband similar to the frequency spectrum of the rectangular pulse of width T seconds in Fig. 8, but shifted up in frequency by f_c c/s.

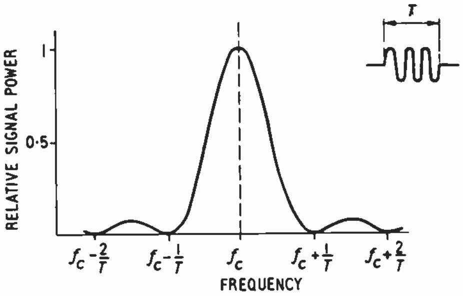


Fig. 9. Frequency spectrum of an amplitude modulated carrier using a binary square wave modulating waveform.

The information carried in each sideband is the same and thus there is 50% redundant information in the two sidebands. The signal carrier itself carries no useful information. Ideally then the information in the signal could be transmitted equally well using only one sideband and thus requiring only half the frequency band. In speech, the lowest useful frequency is about 300 c/s, leaving an appreciable margin between the carrier and each sideband in the signal formed by amplitude modulating a carrier with a speech waveform. Thus a practical filter can remove the carrier and one sideband of this signal, to leave only the other sideband, and so give single sideband suppressed carrier transmission. In data transmission, however, the energy in the modulating waveform is always concentrated towards the very low frequencies, including zero frequency or d.c., and giving rise to the characteristic frequency spectra shown in Fig. 8. Under these conditions it is not practical to use single sideband suppressed carrier transmission. However it is possible to remove a good part of one sideband, and this is done in vestigial sideband a.m. signals, where one sideband, the carrier and only a vestige of the other sideband are transmitted. The carrier itself can also be suppressed to give vestigial sideband suppressed carrier a.m. signals. These are in fact the same as vestigial sideband p.m. signals and are treated separately under p.m. signals in this paper.

One rather unfortunate effect in vestigial sideband signals is the presence of an appreciable quadrature component, resulting from the asymmetry between the two sidebands. The effect of this distortion can be largely eliminated by using synchronous or coherent detection in the receiver, but only at the price of giving the vestigial sideband a.m. system a very low tolerance to frequency modulation effects over telephone circuits. The tolerance of such a system to frequency modulation effects would in fact be similar to that of a p.m. system using coherent detection, and since the latter system could also, when suitably designed, be used with a vestigial sideband signal and with the effective elimination of the quadrature component in the detector, it would of course in general be a much better alternative. Another approach which has often been used is to reduce the depth of modulation in the vestigial sideband a.m. signal to somewhat below 100%, which considerably reduces the level of the quadrature component and enables satisfactory detection to be obtained. The reduction in tolerance to white noise of such an arrangement, resulting from the signal distortion and the means used to reduce it, is however typically about 5 or 6 dB. Such a system has therefore a very low tolerance to both the additive noise and the amplitude modulation effects over telephone circuits.

Again, in a double sideband signal, by virtue of the fact that each sideband carries the same information and since the detected signal in the receiver is derived equally from each sideband, any distortion in one sideband will have appreciably less effect on the receiver detected waveform than if only that one sideband were used to transmit the same information. Indeed one great advantage in using double sideband transmission, and this applies equally to a.m., f.m. and p.m. systems, is a significant increase in tolerance to distortion when compared with the equivalent vestigial sideband systems. Thus where the required speed of working can be achieved satisfactorily using double sideband transmission, this should always be done.

Figure 9 shows the frequency spectrum obtained when a signal carrier of frequency f_c c/s is 100% amplitude modulated by a binary square wave signal in which the element width is T seconds. The signalling speed is therefore $1/T$ bauds or elements per second (assuming here as in the rest of this discussion on transmission rates, that the element width never differs from T seconds). The useful information in each sideband is contained in a frequency band with one boundary at the carrier frequency and having a width of $(1/2T)$ c/s. Thus the bandwidth containing the useful information in a double sideband signal, is $(1/T)$ c/s which is equal to the signalling speed in bauds. Thus the ideal maximum signalling speed for a double sideband a.m. system is one baud per c/s of the frequency band. In the case of a vestigial sideband a.m. system, the maximum signalling speed is about $1\frac{1}{2}$ bauds per c/s of the frequency band, although the ideal maximum would of course approach 2 bauds per c/s. Practical systems are normally used at appreciably lower signalling speeds, preferably not exceeding $\frac{2}{3}$ baud per c/s for a double sideband a.m. system and one baud per c/s for a vestigial sideband a.m. system.

Practical a.m. systems of course do not normally use a rectangular modulating waveform, but more generally one with a "sine^2" or "raised-cosine" shape, thus restricting the frequency band occupied by this waveform as illustrated in Fig. 8. The envelope of each signal element now has a raised-cosine shape and each signal pulse overlaps somewhat into the space allocated to the adjacent pulse on either side. This of course follows the general rule that any tendency to reduce the transmitted frequency band tends to lengthen the transmitted pulses, since for a given pulse shape the width of the frequency spectrum is inversely proportional to the width of the pulse.

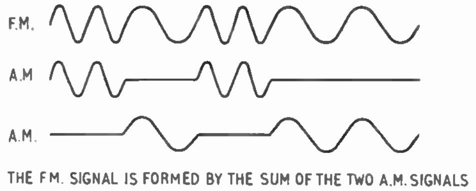

F.M.
A.M
A.M.
THE F.M. SIGNAL IS FORMED BY THE SUM OF THE TWO A.M. SIGNALS

Fig. 10. Relationship between f.m. and a.m. for binary coded digital signals.

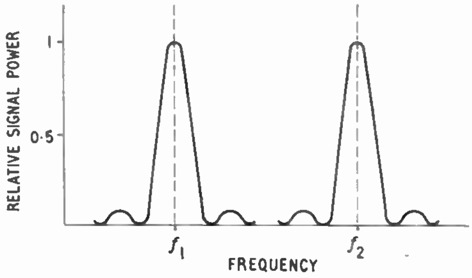
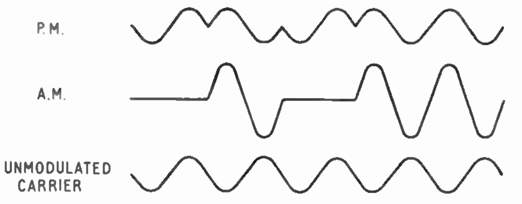


Fig. 11. Frequency spectrum of a frequency modulated carrier using a binary square wave modulating waveform.

4.4. *F.M. Signals*

The relationship between f.m. and a.m. for a binary coded digital signal, using a rectangular modulating waveform, is illustrated in Fig. 10. The f.m. signal may be regarded as being formed by the addition of two separate a.m. signals, one with a carrier frequency of f_1 c/s and the other with a carrier frequency of f_2 c/s. The frequency spectrum is therefore that formed by the addition of the frequency spectra of the two separate a.m. signals. This is shown diagrammatically in Fig. 11. In a practical f.m. signal as used over telephone circuits, of course, the two frequency spectra normally overlap to a considerable degree.

To obtain the optimum performance from such an f.m. system, the frequency separation of the two carrier frequencies (measured in c/s) should have the same value as the signal element rate or the signalling speed in bauds. Under these conditions, assuming the same width for each binary element, the frequency corresponding to the first zero power point below f_2 c/s for the higher frequency spectrum, should be equal to f_1 c/s, and the frequency corresponding to the first zero power point above f_1 c/s for the lower frequency spectrum, should be equal to f_2 c/s. That is, the first inside zero of each spectrum should coincide with the carrier frequency of the other.



THE P.M. SIGNAL IS FORMED BY ADDING THE UNMODULATED CARRIER TO THE A.M. SIGNAL. THIS IS EQUIVALENT TO SUPPRESSING THE CARRIER IN THE A.M. SIGNAL WITHOUT CHANGING THE LEVEL OF THE SIDEBANDS

Fig. 12. Relationship between p.m. and a.m. for binary coded digital signals.

If for a given signalling speed the carrier frequencies are brought closer together, the performance of the f.m. system deteriorates, and in practice the frequency separation (measured in c/s) should not normally be reduced to a value much less than half that of the signalling speed in bauds. A certain reduction in the width of the transmitted signal spectrum can however be achieved in this way.

Thus for an f.m. system the maximum practical signalling speed is around one baud per c/s of the frequency band. Under these conditions each of the two component a.m. signals which make up the f.m. signal, is being used as a vestigial sideband signal and the frequency separation of the two carriers is appreciably less than the optimum. A significant reduction in tolerance to noise and distortion would therefore be experienced by the receiver using this signal. For this reason practical f.m. systems should preferably operate at speeds of no more than about $\frac{1}{2}$ baud per c/s. This compares with $\frac{2}{3}$ baud per c/s for double sideband a.m. systems and one baud per c/s for vestigial sideband a.m. systems.

To obtain the best performance from an f.m. system, the two carrier frequencies should if possible be derived from a single frequency shift oscillator and not from two independent oscillators. The reason for this is that where the two frequencies are derived from one oscillator, there is inevitably a degree of phase coherence between the two different frequencies transmitted and there are no phase discontinuities between adjacent elements, thus narrowing the transmitted signal spectrum and reducing certain distortion effects which may otherwise result in transmission.

A modulating waveform having a somewhat rounded shape may, if required, be used instead of one with a rectangular shape as shown.

4.5. *P.M. Signals*

The relationship between p.m. and a.m. for a binary coded digital signal, using a rectangular modulating waveform, is illustrated in Fig. 12. The p.m. signal may be formed by adding to the a.m. signal, which has twice the peak amplitude of the p.m. signal, an unmodulated carrier having the same frequency as the carrier in the a.m. signal but being 180 deg out of phase and having only half its peak amplitude. The p.m. signal can thus be derived from an a.m. signal of the same peak amplitude, by doubling the amplitude of the a.m. signal and suppressing the carrier. Since the carrier in any a.m. signal contains no useful information, it follows that in tolerance to additive noise a p.m. signal is equivalent to an a.m. signal having twice the peak amplitude of the p.m. signal. This shows very clearly the reason for the basic 6 dB advantage in tolerance to white noise of a p.m. signal over an a.m. signal of the same peak power level. The additional 1 dB advantage of a coherent p.m. system over an a.m. system, is of course because coherent or synchronous detection is used with the p.m. system, whereas envelope detection is normally used with the a.m. system.

As explained above, a p.m. signal is in fact a suppressed carrier a.m. signal and it is also sometimes referred to as bipolar a.m.

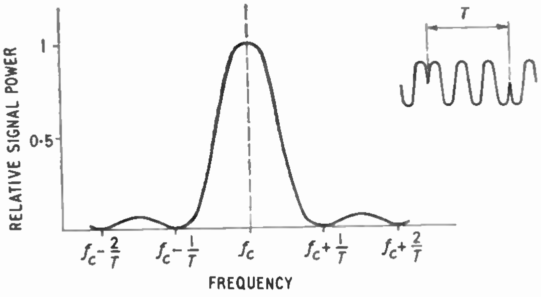


Fig. 13. Frequency spectrum of a phase modulated carrier using a binary square wave modulating waveform.

Because of the close relationship between p.m. and a.m. for binary coded digital signals, the frequency spectrum of a p.m. signal is similar to that of the equivalent a.m. signal, and the spectrum for the p.m. signal is shown in Fig. 13. The p.m. signal can also be transmitted either as a double sideband or as a vestigial sideband signal. Consequently the maximum signalling speed for a p.m. signal is the same as that for the a.m. signal, in the case of either a double sideband or a vestigial sideband signal. Thus the ideal maximum signalling speed for a double sideband p.m. system is one baud per c/s of the frequency band, and the maximum signalling speed for a vestigial sideband p.m. system is about 1½ bauds per c/s of the frequency band, although the ideal maximum would of course approach 2 bauds per c/s. Practical systems should preferably operate at signalling speeds not exceeding ⅔ baud per c/s for a double sideband p.m. system and one baud per c/s for a vestigial sideband p.m. system. This compares with ½ baud per c/s for the equivalent f.m. system.

4.6. *Some Considerations on the Optimum Transmitter Design for P.M. Systems*

In order that the fastest signalling speed over a given bandwidth may be obtained from a practical p.m. system, while still maintaining the most economic arrangement for the transmitter, effective use must be made of one or two simple properties of the waveforms involved. In a p.m. system, the modulating waveform used at the transmitter would normally be either a rectangular waveform or one whose transitions are shaped similarly to those of a raised-cosine pulse. The frequency spectra of these two different waveforms are as shown in Fig. 8.

In a p.m. system where a rectangular modulating waveform is used and where the signal carrier is synchronized to the modulating waveform, when the system is so designed that the transition from one signal element to the next always occurs at a peak of the signal carrier, the larger part of the energy at the modulator output is located at frequencies above the carrier frequency; when the transition from one signal element to the next always occurs at the zero level crossing of the signal carrier, the larger part of the energy is located at frequencies below the carrier frequency. When the transition from one signal element to the next occurs midway between these two points, the frequency spectrum is symmetrical about the carrier. This relationship between the phase of the carrier and the frequency spectrum of the signal thus provides a means of achieving an appreciable filtering action without introducing at the same time any distortion into the modulated signal. It becomes more important as the bit rate of the modulating waveform increases relative to the carrier frequency and for those frequencies further removed from the carrier. It is particularly effective when there are not more than two cycles of the carrier per signal element.

When a rectangular modulating waveform is used and there is no phase coherence between the signal carrier and the modulating waveform at the transmitter, each signal element, depending on the particular phase relationship between the carrier and modulating waveform, contributes more energy to the higher or to the lower frequencies. The transmitted frequency spectrum is now wider and correspondingly more filtering is therefore required at the modulator output to limit the transmitted frequency spectrum to the required frequency band. Because more filtering is required, more delay distortion is introduced into the frequency band, unless yet additional delay equalization is used or unless more complex filters having a linear phase characteristic are used.

For this reason, where a rectangular modulating waveform is used, the signal carrier should always where possible be synchronized or phase locked to the modulating waveform and frequency related to it in such a way that each signal element contains an exact multiple of one half cycle of the carrier with at least one complete cycle. This is of course also the condition for the optimum value of the cross-correlation coefficient.

The alternative approach is to use for the modulating waveform the information-carrying square-wave signal after this has first been passed through a low-pass filter to shape the transitions similarly to those of the raised-cosine pulse in Fig. 8 and giving the waveform therefore a similar frequency spectrum. Provided that the low-pass filter effectively removes all the frequency components in the modulating waveform which are higher than the carrier frequency, there will now be no foldover of the frequency spectrum upon modulation, this being the effect which causes the dependency of this spectrum on the phase relationship between the carrier and modulating waveforms. Consequently a fixed, symmetrical and reasonably narrow frequency spectrum will be obtained for any phase relationship between the signal carrier and modulating waveform. Thus with this arrangement, a carrier frequency which is not synchronized to the modulating waveform can be used with no widening of the frequency spectrum and therefore no need for further filtering. The arrangement has however the disadvantage that a linear double-balanced modulator must be used in place of the simpler balanced gate circuit needed for the other arrangement. Also, in those applications where the carrier frequency is synchronized to the modulating waveform, all the shaping of the transmitted frequency spectrum must now be achieved in the normal manner using filters and the system loses the versatility of the other arrangement, whereby a simple change in the phase relationship between the carrier and modulating waveform can be used to give an appreciable shift in the frequency location of the transmitted energy.

Thus in applications where the carrier is synchronized to the modulating waveform, the former arrangement using a rectangular modulating waveform is probably the best, whereas in applications where the carrier is not synchronized to the modulating waveform, the latter arrangement using a suitably rounded modulating waveform is to be preferred. Where possible the carrier should of course always be synchronized to the modulating waveform.

4.7. *Comparison of A.M., F.M. and P.M. Systems*

Both a.m. and p.m. systems are appreciably more efficient as far as bandwidth is concerned than are f.m. systems, and this is therefore an additional advantage of p.m. systems when compared with f.m. systems. A p.m. system using coherent or differentially-coherent detection thus not only has a greater tolerance to additive noise than the equivalent f.m. system, but it is also capable of a higher signalling speed.

As the signalling speed increases from ½ to 1 baud per c/s of the frequency band, so the tolerance to noise of an f.m. system inevitably becomes further reduced relative to that of a p.m. system. Under these conditions a p.m. system using differentially-coherent detection has an advantage in tolerance to white noise, over the equivalent f.m. system, of more than 3 dB, at typical error rates of around 1 bit in 10^5.

A p.m. system using differentially-coherent detection will therefore in practice give a considerably better overall performance over telephone circuits than the equivalent f.m. system.

5. Transmission of Timing Information in Complete Data Transmission Systems

5.1. *The Need for Timing Information*

A binary coded digital signal on its own has very little meaning in the absence of the corresponding timing information. Unless the duration or width of a signal element is known, there is no means of telling how many "1's" and "0's" are contained in any part of the signal waveform. Thus the receiver which is used to detect any information carrying signal must always have a means of extracting from this signal the transmitted element rate and of producing this information in the form of a timing waveform. Whether this is done in the receiver itself or in the associated equipment, it is always absolutely essential that this be done, because without it the correct interpretation of the received information is not possible.

The transmission of timing information may be achieved in two basically different ways. The first of these is found in synchronous data transmission systems and the second in start-stop systems. The operation of these two methods is illustrated in Fig. 14. As before, only binary coded systems will be considered here.

5.2. *Synchronous Systems*

The transmission of timing information in synchronous systems may itself be achieved in two different ways. In the first of these the timing information is transmitted by a completely separate signal, which may use a separate modulated carrier or may be imposed on the signal carrier itself, using a different modulation method to that used for the signal. Sometimes even the same modulation method may be used. The essential requirement of this or any other method is of course that the timing waveform obtained in the receiver should not be affected by the frequency modulation effects over telephone circuits, and for this reason a single transmitted frequency on its own cannot be used.

The above method has the advantage that the correct transmission of the timing information is in no way dependent on the signal code or the transmitted signal pattern. It has however a number of disadvantages. Firstly, because an additional signal is transmitted there is an inevitable increase in the complexity of both the transmitter and the receiver. Secondly, because over telephone circuits a limit is set on the peak transmitted signal power, the presence of the additional signal carrying the timing information necessarily implies a lower level for the signal carrying the transmitted message information and therefore a lower tolerance of the latter signal to the additive noise over telephone circuits. Thirdly, there may sometimes be a degree of interference between the signal carrying the timing information and that carrying the transmitted message information, in some of those applications where the two frequency spectra overlap, thus reducing the tolerance of the system to both noise and distortion. Fourthly, in the case of f.m. and p.m. systems it may sometimes be necessary to use a.g.c. in the receiver which could otherwise slice the received signal, and this inevitably introduces the various disadvantages associated with a.g.c. Finally, and where it applies perhaps of most importance, in those cases where the frequency spectrum of the signal carrying the timing information does not occupy quite the same parts of the available frequency band as does that of the signal carrying the transmitted message information, for the corresponding modulating frequencies, the phase of the timing waveform obtained in the receiver relative to that of the detected signal, is to some degree dependent on the group-delay-frequency characteristic of the transmission path. Thus the timing waveform at the receiver must always be correctly re-phased relative to the detected signal for any new transmission path having a different group-delay-frequency characteristic or for any significant change in the group-delay-frequency characteristic of the path used. However, one important requirement for a data transmission system is that once the equipment has been adjusted it should be capable of operating correctly over any one of possibly a large group of different telephone circuits, without further readjustment. Such arrangements for transmitting the timing information, where the phase of the timing waveform at the receiver relative to that of the detected signal is a function of the group-delay-frequency characteristic of the transmission path, cannot therefore be used where it is essential that these requirements be satisfied.

Because of these various disadvantages, this method of transmitting the timing information should only be used in those cases where the system must be able to handle the more unfavourable signal patterns, such as those containing frequent unbroken sequences of "1's" or "0's" each exceeding say 50 bits in length. In most normal applications, particularly where error detecting or correcting codes are used, it is very unlikely and often impossible for such signal patterns to be obtained. For these applications the alternative method of transmitting the timing information is to be preferred. In this method, the signal carrying the transmitted message information is also used to carry the timing information and for this reason the method relies on the signal pattern transmitted. However, since the method does not require an additional signal to carry the timing information, it does not suffer from any of the disadvantages mentioned for the other method. One way in which the signal carrying the transmitted message information may also be used to carry the timing information is outlined as follows.

The detected signal in the receiver, for a binary coded system, has two different voltage levels, one representing "1" and the other "0". The transition between a "1" and a "0" or vice versa always occurs at the junction between the two adjacent detected signal elements and therefore if the received signal element rate is known at the receiver, these transitions can be used to phase correctly a timing waveform of the correct frequency which is already generated by an oscillator, or alternatively the transitions may be used themselves both to generate and correctly phase the timing waveform. The latter method may for instance be achieved in practice as illustrated in Fig. 14. Each transition in the detected waveform is used to generate a positive going pulse. The waveform comprising these pulses has a very strong frequency component not only at the correct signal element rate but also in the correct phase. This frequency component is filtered out using a narrow band filter and it is then suitably shaped to produce the timing waveform. This, as can be seen from the diagram, is phased so that the positive-going pulses sample each detected signal element centrally to produce the binary square wave output signal. The particular method of extracting the timing waveform from the detected signal is of course in detail only one of the many quite different approaches, but in principle it follows the same general pattern which must always be used. This is that the timing waveform must in some way be extracted from the received and preferably detected signal itself (or its equivalent) and then used to sample the detected signal at the optimum points in the waveform, thus producing an undistorted binary square wave output signal which together with the timing waveform provide the required information at the receiver output.

5.3. *Start-Stop Systems*

The alternative basic approach in the transmission of timing information is that used in start-stop systems. Here the timing information is carried by one or more special signal elements which are interspersed at regular intervals between the signal elements carrying the transmitted message information. In other words, special elements in the transmitted information are allocated solely to the transmission of the timing information. The example in

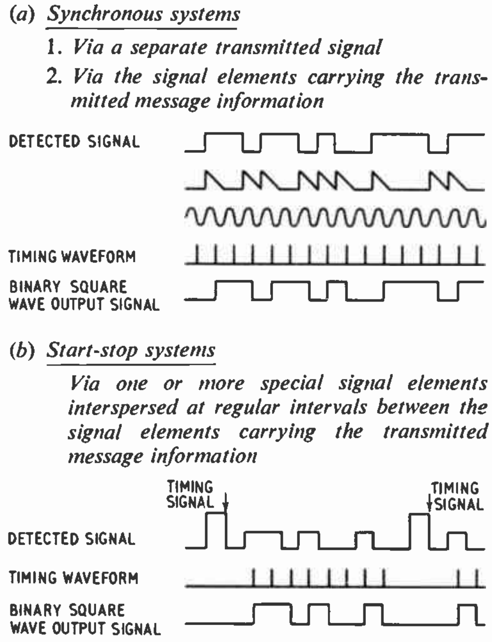

(a) Synchronous systems
1. Via a separate transmitted signal
2. Via the signal elements carrying the transmitted message information
DETECTED SIGNAL
TIMING WAVEFORM
BINARY SQUARE WAVE OUTPUT SIGNAL
(b) Start-stop systems
Via one or more special signal elements interspersed at regular intervals between the signal elements carrying the transmitted message information
TIMING SIGNAL
TIMING SIGNAL
DETECTED SIGNAL
TIMING WAVEFORM
BINARY SQUARE WAVE OUTPUT SIGNAL

Fig. 14. Transmission of timing information.

Fig. 14 shows an arbitrary detected signal, in which the detected elements carrying the timing information have twice the amplitude of the elements carrying the transmitted message information. The timing signals could alternatively have the same amplitude or form as the message signals but be differentiated from these through the use of some unique code. Various other ways can of course also be used to enable the receiver to differentiate between the signals carrying timing information and those carrying the transmitted message information. At the receiver, each timing signal is used to generate a discrete series of timing pulses, which are used to sample the following group of detected signal elements and thus provide at the receiver output the information carrying binary square wave signal together with the required timing waveform.

5.4. *Comparison of Synchronous and Start-Stop Systems*

The advantages of start-stop systems are briefly as follows. Since the timing information is transmitted by special signal elements, immediately the receiver has received one of these it becomes fully operative with therefore a negligible delay in starting and no special start or synchronizing characters required in the transmitted code. A synchronous system on the other hand requires a certain minimum time at the beginning of each transmission for the timing waveform generator to become fully operative, although with correct design this time need not in practice exceed 10 or 20 milliseconds. Also a special start code is required at the beginning of each transmission to enable the equipment fed from the receiver to identify the first, second, third, etc. element in each character. A start-stop system therefore requires less delay before the receiver is fully operative, and because no special synchronizing or start characters are required in the code, the equipment may sometimes be less complex than that for the equivalent synchronous system.

There are however two rather serious disadvantages associated with start-stop systems. Firstly, because the timing information is carried by special signal elements, which cannot therefore be used at the same time to carry message information, the effective or useful transmission rate is reduced relative to that obtainable with the equivalent synchronous system. Although the latter requires a special start code at the beginning of each transmission, since the start code need only be one or two characters in length and the transmitted message may contain hundreds or even thousands of characters, the reduction in the effective rate of transmission is generally in practice quite negligible. Secondly, because in a start-stop system the timing waveform used to sample the detected signal is for each character derived from a single timing element at the beginning of that character, the phase of the timing waveform used for each character is determined by only the one signal element. Hence any noise or distortion in this signal element affects the position of the following timing waveform to the same degree as it affects the timing signal element itself. Consequently the timing waveform in any start-stop system is much more seriously affected by both noise and distortion than it is in any synchronous system, where the timing waveform is derived by a process of integration over many signal elements. Start-stop systems therefore inevitably have a significantly lower tolerance to both noise and distortion.

On account of these rather serious disadvantages of start-stop systems, synchronous data transmission systems should always be used in preference whenever possible.

5.5. *Complete Synchronous and Start-Stop Systems*

In a synchronous system in which the carrier is synchronized to the modulating waveform at the transmitter, the signal carrier and a timing waveform synchronized to this carrier are usually both generated from the same oscillator in the transmitter. The timing waveform is fed to the associated equipment in which it is used to synchronize the information-carrying square wave signal which is fed from this equipment to the transmitter. Alternatively the transmitter may simply be used to generate a carrier from a timing waveform received from the associated equipment, this timing waveform being synchronized to the information-carrying square wave signal which is also fed to the transmitter from the associated equipment. This latter arrangement would normally only be used in those applications where the equipment associated with the transmitter must work at its own speed and cannot be slaved to an external timing waveform as in the previous arrangement.

In a synchronous system in which the carrier is not synchronized to the modulating waveform in the transmitter, the signal carrier is generated by an oscillator in the transmitter, and the information carrying square wave signal, which is fed to the transmitter from the associated equipment, is synchronized to a separate frequency source in this equipment. It is therefore not synchronized or phase related in any way to the signal carrier, and no timing waveform is fed either to or from the transmitter.

In a start-stop system, because the modulation rate is determined essentially by the equipment associated with the transmitter and not by the transmitter itself, and because the carrier is not normally synchronized to the modulating waveform, no timing waveform is fed either to or from the transmitter.

The basic requirements for the receiver are the same in any system. This is because the receiver is always required to provide at its output firstly a binary square wave signal, which should be the same as that fed to the transmitter at the other end of the line, with the possible exception of the timing elements in a start-stop system, and secondly it must provide the associated timing waveform, without which the information in the binary square wave signal cannot be correctly interpreted. For this reason of course there is no such thing as an asynchronous data transmission system which is at the same time a complete system. Any receiver, which gives no timing waveform at its output, is necessarily incomplete and the required timing information must in that case always be obtained in the associated equipment.

Another important point which is sometimes overlooked, is that in all data transmission systems, with the sole exception of the synchronous system in which the timing information is transmitted via a separate signal, the receiver must have a built-in knowledge of the modulation rate of the received signal. This knowledge, in the case of a synchronous system in which the timing information is carried by the message information signal itself, is usually contained in the form of a tuned circuit or narrow band filter which extracts the required frequency component from a waveform derived from the detected signal; alternatively it is contained in the form of a phase-locked oscillator whose phase is controlled by this waveform. In the case of a start-stop system, a start-stop oscillator, started and phased by the timing signals, is required to provide the modulation rate information. In the synchronous system in which the timing information is transmitted via a separate signal, the timing waveform at the receiver output is normally obtained after first filtering the received and detected timing signal, to reduce as far as possible the effects of both noise and distortion. If no filtering is used in such a system, the tolerance to noise and distortion are not significantly better than that of a typical start-stop system. For these reasons it is therefore true to say that in any system the receiver would normally always have a built-in knowledge of the modulation rate of the received signal.

Therefore not only is there no such thing as an asynchronous data transmission system which is at the same time a complete system, but also in any complete system the receiver or its associated equipment must normally always have a built-in knowledge of the signal modulation rate. Thus in a complete data transmission system, if this modulation rate is changed, correct operation can only be maintained if a corresponding change is also made in the modulation rate for which the receiver or its associated equipment is adjusted. This very important point is often overlooked when different modulation methods are compared on the basis of their adaptability for different applications over telephone circuits. Whatever modulation method is used, the receiver will in practice only work correctly at one modulation rate, and therefore although the correct operation of some modulation methods is more dependent on the modulation rate than that of others, this is seldom in practice of any importance at all when the design of a complete data transmission system is being considered.

5.6. *Asynchronous Systems*

Where it is required to transmit at any one of a number of different signalling speeds, using as much common transmission equipment as possible for the different speeds, a very different situation is created. By the very nature of the requirement, an appreciable degradation in performance must be accepted at any one signalling speed relative to that which could be obtained with the optimum data transmission system specifically designed for that speed of working. One of the reasons for this is, of course, that less information can either be contained in the transmitted waveform or extracted from it by the receiver, because neither the transmitter modulator nor the receiver detector can have any knowledge of the signal modulation rate.

One solution in this case is to use, for the common equipment, an f.m. system in which the complete timing waveform generator in the receiver is omitted. No attempt would also be made at the transmitter to synchronize the signal carrier to the modulating waveform and therefore no timing waveform would be required here either. Under these conditions the common transmission equipment accepts a binary square wave signal at the transmitter, converts it to the corresponding frequency modulated carrier signal, which at the receiver is converted back again into the original binary square wave signal, although with possibly a certain amount of distortion. Such a system is sometimes referred to as an asynchronous system, containing as it does only the asynchronous part of a complete data transmission system and not being therefore a complete system on its own.

Alternatively a p.m. system using coherent detection may be used, in which the ambiguity in the binary value associated with the phase of the received signal carrier is overcome by sending at the beginning of each transmission a continuous series of "0's" or "1's" or some other convenient code, whereby the receiver can determine the carrier phase associated with either a "0" or a "1". The receiver can then remember by suitable means the phase of the signal carrier which corresponds to either a "0" or a "1" at any point during the rest of the transmission. Such an arrangement can be made to work satisfactorily so long as the phase variations in the carrier resulting from excursions in the frequency location of the received signal spectrum are not rapid enough to be confused with a real transmitted phase inversion between adjacent elements. Since such a frequency excursion would in any case always result in an error, this requirement is not a serious limitation of the system. As in the case of the asynchronous f.m. system, the complete timing waveform generator in the receiver would be omitted and no attempt would be made at the transmitter to synchronize the signal carrier to the modulating waveform, no timing waveform being therefore needed here either.

Again there are other arrangements of a p.m. system using coherent detection, in which the receiver is enabled by some suitable means to identify the phase of the received signal carrier corresponding to a "0" or a "1" at any point during transmission. These are however in general either more complex, less efficient or less versatile than the particular arrangement outlined above, and they are therefore probably less suitable for this application.

The disadvantages of either of the two systems considered above are that an additional timing waveform generator of some kind or another must always be used with the receiver and, of considerably more importance perhaps, in many applications an inferior data transmission system must be used. In addition, because any such system must obviously be designed for the highest of the range of signalling speeds over which it will be used, and because its tolerance to noise is therefore effectively limited to that of the highest signalling speed, when used at any lower signalling speed there is yet a further reduction in tolerance to noise relative to that which would be obtained using the optimum system specifically designed for that signalling speed. For the lower signalling speeds, the total degradation in performance can obviously become quite serious.

The main use for asynchronous systems is in those applications where it is required to standardize the data transmission equipment and where the peripheral equipment from which it is required to transmit data, can only work at a fixed speed which differs from one type of equipment to another. For these applications an asynchronous system of either type outlined above could be used.

In practice, however, a large number of the types of peripheral equipment between which it is required to transmit data are of the type which enable the transmitter to dictate the signalling speed. With all these types of equipment it would be possible and indeed highly desirable to standardize on a fixed speed data transmission system, which is designed to give both the optimum tolerance to noise and distortion and the maximum signalling speed for the class of telephone circuits to be used. Where necessary, different standard signalling speeds could be used for different types or grades of telephone circuits, each standard speed being the optimum for its particular grade of circuit. In this way each type of peripheral equipment will always obtain the best possible use of the telephone circuits and with no necessary increase in the complexity of the data transmission equipment, since there is no very significant difference in the complexity of the equipment required for an f.m. system or a p.m. system using either coherent or differentially-coherent detection.

6. Conclusions

For any application over telephone circuits, a synchronous data transmission system will give a better performance than a start-stop system.

Where permitted by the signalling speed required, double sideband transmission should be used with any type of modulation.

Either a p.m. or an a.m. system enables a higher signalling speed to be obtained over a given bandwidth than an f.m. system.

In tolerance to additive noise, a p.m. system using coherent detection gives the best performance closely followed by a p.m. system using differentially-coherent detection. The latter is somewhat better than an f.m. system which in turn is better than an a.m. system. An a.m. system has a very much lower tolerance than the other systems to amplitude modulation effects, and a p.m. system using coherent detection has a very much lower tolerance than the other systems to frequency modulation effects. Thus the best overall tolerance to the various types of additive and multiplicative noise over telephone circuits is obtained by a p.m. system using differentially-coherent detection.

In applications where it is required to transmit at any one of a number of different signalling speeds, using as much common transmission equipment as possible, either an asynchronous f.m. system or an asynchronous p.m. system using coherent detection, may be used. As these are not complete systems, a timing waveform generator suitably designed for the particular signalling speed required must in every case be added to the receiver. However the tolerance to noise of such an arrangement may in many applications be considerably below that obtainable with the optimum data transmission system.

Over telephone circuits which do not contain carrier links and over which frequency modulation effects do not therefore occur and also over those telephone carrier circuits on which the maximum shift in the frequency location of the received signal spectrum is strictly limited to one or two cycles per second, a p.m. system using coherent detection is probably the best system, provided that a double sideband signal is used and that no-signal is not used as a standby condition.

The best average performance over all types of telephone circuits and in most general applications, will however most probably be given by a p.m. system using differentially-coherent detection.

7. Acknowledgment

The author is indebted to the Director of Research, British Telecommunications Research Ltd., for permission to publish this paper.

8. Bibliography

These references are some of the more important and readily available of the contributions whose opinions, results and conclusions have been considered in the preparation of this paper. Their titles are here grouped together as far as possible according to their subject matter and are arranged in seven different sections. The contributions or papers listed in Section 8.1 are largely mathematical and are concerned with the more basic theory. The papers in Section 8.2, although partly theoretical, are in general of a more practical nature, whereas those in the remaining sections are concerned mainly with the description and assessment of different practical data transmission systems.

8.1. *Theoretical Analyses of the Performance and Particular Properties of Different Modulation Methods*

1. H. Nyquist, "Certain factors affecting telegraph speed", *Bell Syst. Tech. J.*, **3**, pp. 324–46, April 1924.
2. H. Nyquist, "Certain topics in telegraph transmission theory", *Trans. Amer. Inst. Elect. Engrs*, **47**, pp. 617–44, April 1928.
3. R. V. L. Hartley, "Transmission of information", *Bell Syst. Tech. J.*, **7**, pp. 535–63, July 1928.
4. D. Gabor, "Theory of communication", *J. Instn Elect. Engrs*, **93**, Part III, No. 26, pp. 429–57, November 1946.
5. C. E. Shannon, "A mathematical theory of communication", *Bell Syst. Tech. J.*, **27**, Nos. 3 and 4, pp. 379–423 and pp. 623–56, July and October 1948.
6. J. R. Davey and A. L. Matte, "Frequency shift telegraphy —radio and wire applications", *Bell Syst. Tech. J.*, **27**, No. 2, pp. 265–303, April 1948.
7. S. Reiger, "Error probabilities of binary data transmission systems in the presence of random noise", *Convention Record of the I.R.E.*, 1953, Part 8, pp. 72–9.
8. G. F. Montgomery, "A comparison of amplitude and angle modulation for narrow-band communication of binary-coded messages in fluctuation noise", *Proc. Inst. Radio Engrs*, **42**, pp. 447–54, February 1954.
9. D. Middleton, "Statistical theory of signal detection". *Trans. Inst. Radio Engrs (Information Theory)*, No. PGIT–3, pp. 26–51, March 1954.
10. R. C. Davis, "The detectability of random signals in the presence of noise", *Trans. Inst. Radio Engrs (Information Theory)*, No. PGIT–4, pp. 52–62, March 1954.
11. W. W. Peterson, T. G. Birdsall and W. C. Fox, "The theory of signal detectability", *Trans. Inst. Radio Engrs (Information Theory)*, No. PGIT–4, pp. 171–212, September 1954.
12. W. R. Bennett, "Methods of solving noise problems", *Proc. Inst. Radio Engrs*, **44**, pp. 609–38, May 1956.
13. S. Goldman, "Information Theory", pp. 235–41 (Prentice-Hall, New York, 1955).
14. M. Schwartz, "Information, Transmission, Modulation and Noise", pp. 423–31 (McGraw-Hill, New York, 1959).
15. P. Mertz, "Information theory impact on modern communications", *Communication and Electronics*, No. 32, pp. 431–7, September 1957.
16. E. D. Sunde, "Theoretical fundamentals on pulse transmission", *Bell Syst. Tech. J.*, **33**, pp. 721–88 and 987–1010, May and July 1954.
17. E. D. Sunde, "Ideal binary pulse transmission by a.m. and f.m.", *Bell Syst. Tech. J.*, **38**, No. 6, pp. 1357–1426, November 1959.
18. E. D. Sunde, "Pulse transmission by a.m., f.m. and p.m. in the presence of phase distortion", *Bell Syst. Tech. J.*, **40**, No. 2, pp. 353–422, March 1961.
19. M. Masonson, "Binary transmissions through noise fading", *I.R.E. National Convention Record*, Part 2, pp. 69–82, 1957.
20. G. L. Turin, "Error probabilities for binary symmetric ideal reception through nonselective slow fading and noise", *Proc. Inst. Radio Engrs*, **46**, pp. 1603–19, September 1958.
21. J. G. Lawton, "Theoretical error rates of 'differentially coherent' binary and 'kineplex' data transmission systems", *Proc. Inst. Radio Engrs*, **47**, pp. 333–4, February 1959.
22. C. R. Cahn, "Performance of digital phase-modulation communication systems", *Trans. Inst. Radio Engrs (Communications Systems)*, **CS-7**, No. 1, pp. 3–6, May 1959.
23. A. B. Glenn, "Performance analysis of a data link system", *Trans. Inst. Radio Engrs (Communications Systems)*, **CS-7**, No. 1, pp. 14–24, May 1959.
24. A. B. Glenn, "Comparison of psk *vs* fsk and psk-am *vs* fsk-am binary coded transmission systems", *Trans. Inst. Radio Engrs (Communications Systems)*, **CS-8**, No. 2, pp. 87–100, June 1960.
25. C. R. Cahn, "Combined digital phase and amplitude modulation communication systems", *Trans. Inst. Radio Engrs (Communications Systems)*, **CS-8**, No. 3, pp. 150–5, September 1960.
26. J. C. Hancock and R. W. Lucky, "Performance of combined amplitude and phase-modulated communication systems", *Trans. Inst. Radio Engrs (Communications Systems)*, **CS-8**, No. 4, pp. 232–7, December 1960.
27. C. W. Helstrom, "The comparison of digital communication systems", *Trans. Inst. Radio Engrs (Communications Systems)*, **CS-8**, No. 3, pp. 141–50, September 1960.
28. R. Filipowski, "Recent progress in applying information theory to digital transmission systems", *Communication and Electronics*, No. 40, pp. 848–55, January 1959.
29. D. G. Tucker, "Synchronous demodulation of phase-reversing binary signals, and the effect of limiting action" *Trans. Inst. Radio Engrs (Communications Systems)*, **CS-9**, No. 1, pp. 77–82, March 1961.

30. N. B. Bobrov, "The transmission of discrete signals by means of phase difference modulation", *Telecommunications (U.S.S.R.)*, No. 6, pp. 599–608, 1960.

31. H. F. Harmuth, "Transmission of information by orthogonal time functions", *Communication and Electronics*, No. 49, pp. 248–55, July 1960.

8.2. *Noise and Distortion over Telephone Circuits and their Effects on Data Transmission Systems*

32. "Facilities for data transmission over Post Office links", *British Communications and Electronics*, **6**, No. 2, pp. 120–2, February 1959.

33. "Facilities for Data Transmission", booklet issued by H.M. Postmaster General.

34. B. R. Horsfield and D. C. Smith, "A.C. Signalling—A Review of Current Problems", The Institution of Post Office Electrical Engineers Printed Paper, No. 204, pp. 4–7, April 1952.

35. T. Combellick, "Synchronization of single-sideband carrier systems for high speed data transmission", *Trans. Inst. Radio Engrs (Communications Systems)*, **CS-7**, No. 2, pp. 110–4, June 1959.

36. M. B. Williams, "Present and future facilities for data transmission", *Computer J.*, **4**, No. 2, pp. 88–95, July 1961.

37. "Characteristics of P.O. Tariff-E Private Wires for Data Transmission", P.O. Engineering Department, LMD Branch, Technical Memorandum, No. 6000, Issue 1, February 1960.

38. "Performance Specification for Customers' Data Transmission Devices for use on the Public Telephone System", Draft Specification from the P.O. Engineering Department, Subscriber Apparatus Branch, December 1960.

39. P. Mertz, "Transmission line characteristics and effects on pulse transmission", "Proceedings of the Symposium on Information Networks", Vol. 3, pp. 85–114 (Polytechnic Institute of Brooklyn, April 1954).

40. P. Mertz, "The effect of delay distortion on data transmission", *Communication and Electronics*, No. 49, pp. 228–32, July 1960.

41. A. D. Fowler and R. A. Gibby, "Assessment of effects of delay distortion in data systems", *Communication and Electronics*, No. 40, pp. 918–23, January 1959.

42. R. A. Gibby, "An evaluation of a.m. data system performance by computer simulation", *Bell Syst. Tech. J.*, **39**, No. 3, pp. 675–704, May 1960.

43. R. A. Gibby, "An Evaluation of F.M. Data System Performance by Computer Simulation", Paper presented at the International Symposium on Data Transmission, Delft, September 1960.

44. W. R. Bennett and F. E. Froehlich, "Techniques for Comparing Modulation Methods for Data Transmission over Telephone Circuits", Paper presented at the International Symposium on Data Transmission, Delft, September 1960.

45. P. Mertz, "Model of impulsive noise for data transmission", *Trans. Inst. Radio Engrs (Communications Systems)*, **CS-9**, pp. 130–7, June 1961.

46. A. A. Alexander, R. M. Gryb and D. W. Nast, "Capabilities of the telephone network for data transmission", *Bell Syst. Tech. J.*, **39**, No. 3, pp. 431–76, May 1960.

47. "Delay distortion", *Lenkurt Demodulator*, **9**, No. 7, pp. 1–11, July 1960.

48. "Noise", *Lenkurt Demodulator*, **9**, No. 4, pp. 1–7, April 1960.

49. "Crosstalk", *Lenkurt Demodulator*, **9**, No. 11, pp. 1–11, November 1960.

50. R. G. Enticknap, "Errors in data-transmission systems", *Trans. Inst. Radio Engrs (Communications Systems)*, **CS-9**, No. 1, pp. 15–20, March 1961.

51. E. P. G. Wright, "Error rates and error distributions on data transmitted over switched telephone connections", *Trans. Inst. Radio Engrs (Communications Systems)*. **CS-9**, No. 1, pp. 12–15, March 1961.

52. G. A. Wildhagen, "Some Results of Data Transmission Tests over Long Telephone Circuits", Paper presented at the International Symposium on Data Transmission, Delft, September 1960.

53. "A.m. versus f.m. for digital data transmission", *Lenkurt Demodulator*, **7**, No. 2, pp. 1–7, February 1958.

54. "Vestigial sidebands in high speed data transmission", *Lenkurt Demodulator*, **7**, Nos. 8–9, pp. 1–6, August–September 1958.

55. "Faster data transmission", *Lenkurt Demodulator*, **9**, No. 2, pp. 1–7, February 1960.

56. "Methods of transmitting data faster", *Lenkurt Demodulator*, **9**, No. 3, pp. 1–9, March 1960.

8.3. *A.M. Systems*

57. C. A. Lovell, J. H. McGuigan and O. J. Murphy, "An experimental polytonic signalling system", *Bell Syst. Tech. J.*, **34**, No. 4, pp. 783–806, July 1955.

58. A. W. Horton and H. E. Vaughan, "Transmission of digital information over telephone circuits", *Bell Syst. Tech. J.*, **34**, No. 3, pp. 511–28, May 1955.

59. C. R. Doty and L. A. Tate, "A data transmission machine", *Communication and Electronics*, No. 27, pp. 600–3, November 1956.

60. A. P. Clark, "A high speed signalling system for use over telephone circuits", *A.T.E. Journal*, **15**, No. 2, pp. 157–72, April 1959.

61. A. P. Clark, "A High Speed Data Transmission System for use over Telephone Circuits", AGARDograph 43, pp. 111–139. Paper presented at the Data Handling Meeting, AGARD Avionics Panel, Aachen, Germany. September 1959.

62. A. Girinsky and P. Roussel, "High speed transmission of numerical data over telephone channels", *Electrical Communication*, **36**, No. 4, pp. 248–62, 1960.

63. J. V. Harrington, P. Rosen and D. A. Spaeth, "Some results on the transmission of pulses over lines", "Proceedings of the Symposium on Information Networks", Vol. 3, pp. 115–30 (Polytechnic Institute of Brooklyn, April 1954).

64. R. G. Enticknap and E. F. Schuster, "SAGE data system considerations", *Communication and Electronics*. No. 40, pp. 824–32, January 1959.

65. R. O. Soffel and E. G. Spack, "SAGE data terminals", *Communication and Electronics*, No. 40, pp. 872–9, January 1959.

66. R. T. James, "Communication channels for SAGE data systems", *Communication and Electronics*, No. **40**, pp. 838–43, January 1959.

67. A. E. Ruppel, "Sage data transmission service", *Bell Laboratories Record*, **35**, No. 10, pp. 401–405, October 1957.

68. E. A. Irland, "A high speed data signaling system", *Bell Laboratories Record*, **36**, No. 10, pp. 376–80, October 1958.

69. J. L. Hollis, "Digital data fundamentals and the two level vestigial sideband system for voice bandwidth circuits", *I.R.E. Wescon Convention Record*, Part 5, pp. 132–45, 1960.

70. J. L. Hollis, "Sending digital data over narrow-band lines", *Electronics*, **32**, No. 23, pp. 72–4, 5th June 1959.

71. G. Holland and J. C. Myrick, "A 2500 baud time-sequential transmission system for voice frequency wire line transmission", *I.R.E. National Convention Record*, Part 8, pp. 187–90, 1959.

72. G. Holland and J. C. Myrick, "A 2500 baud time-sequential transmission system for voice frequency wire line transmission", *Trans. Inst. Radio Engrs (Communications Systems)*, **CS-7**, No. 3, pp. 180–4, September 1959.

73. J. L. Hollis, "Measured performance of the Sebit-25 data system over wire line facilities at 2500 bits per second", *Trans. Inst. Radio Engrs (Communications Systems)*, **CS-8**, No. 2, pp. 134–7, June 1960.

74. J. Labeyrie, "High speed data transmission over a group link telephone channel", *Trans. Inst. Radio Engrs (Communications Systems)*, **CS-9**, No. 1, pp. 66–9, March 1961.

8.4. *F.M. Systems*

75. W. A. Malthaner, "Experimental data transmission system", *I.R.E. Wescon Convention Record*, Part 8, pp. 56–63, 1957.

76. R. M. Gryb, " 'Recorded carrier' system for high speed data transmission", *Bell Laboratories Record*, **35**, No. 9, pp. 321–5, September 1957.

77. A. Boggs and J. E. Boughtwood, "Application of telegraph techniques in data transmission", *Western Union Tech. Rev.*, **13**, No. 3, pp. 90–7, July 1959.

78. A. Boggs and J. E. Boughtwood, "Application of telegraph techniques in data transmission", *Communication and Electronics*, No. 44, pp. 336–40, September 1959.

79. L. A. Weber, "A frequency modulation digital subset for data transmission over telephone lines", *Communication and Electronics*, No. 40, pp. 867–72, January 1959.

80. "Outline Specification for Post Office 1000 bauds Serial Data Transmission Set", Draft Specification No. 4 from the P.O. Engineering Department, LMD Branch.

81. M. L. Doelz, "Predicted-wave radio teleprinter", *Electronics*, **27**, No. 12, pp. 166–9, December 1954.

8.5. *P.M. Systems*

82. J. P. Costas, "Phase-shift radio teletype", *Proc. Inst. Radio Engrs*, **45**, pp. 16–20, January 1957.

83. P. A. Chittenden, "Notes on transmission of data at 750 bauds over practical circuits", *Trans. Inst. Radio Engrs (Communications Systems)*, **CS-9**, No. 1, pp. 7–12, March 1961.

84. E. Hopner, "An experimental modulation-demodulation scheme for high-speed data transmission", *I.B.M. J. Res. Devel.*, **3**, No. 1, pp. 74–84, January 1959.

85. E. Hopner, "Phase reversal data transmission system for switched and private telephone line applications", *I.B.M. J. Res. Devel.*, **5**, No. 2, pp. 93–105, April 1961.

86. "I.B.M. Data Transmission Tests", Paper presented at the International Symposium on Data Transmission, Delft, September 1960.

87. J. R. Masek, "Carrier phase reversal transmits digital data over telephone lines", *Electronics*, **34**, No. 21, pp. 56–8, 26th May 1961.

88. M. L. Doelz, E. T. Heald and D. L. Martin, "Binary data transmission techniques for linear systems", *Proc. Inst. Radio Engrs*, **45**, pp. 656–61, May 1957.

89. R. R. Mosier and R. G. Clabaugh, "Kineplex, a bandwidth-efficient binary transmission system", *Communication and Electronics*, No. 76, pp. 723–8, January 1958.

8.6. *Transmission of Timing Information*

90. J. O. Edson, M. A. Flavin and A. D. Perry, "Synchronized clocks for data transmission", *Communication and Electronics*, No. 40, pp. 832–6, January 1959.

8.7. *Comparison of Different Systems*

91. P. Mertz and D. Mitchell, "Transmission aspects of data transmission service using private line voice telephone channels", *Bell Syst. Tech. J.*, **36**, No. 6, pp. 1451–86, November 1957.

92. J. V. Beard and A. J. Wheeldon, "A comparison between alternative h.f. telegraph systems", *Point to Point Telecommunications*, **4**, No. 3, pp. 20–48, June 1960.

93. J. M. Wier, "Digital data communication techniques", *Proc. Inst. Radio Engrs*, **49**, pp. 196–209, January 1961.

94. "A Report on High-Speed Communications Equipment", U.S. Federal Aviation Agency, pp. 1–173, 1959.

Manuscript first received by the Institution on 5th October 1961, and in final form on 22nd December 1961 (Paper No. 724).

The discussion which followed the presentation of this paper at the Symposium will be published in a later issue of the *Journal*.

APPLICANTS FOR ELECTION AND TRANSFER

As a result of its meeting on 3rd May the Membership Committee recommended to the Council the following elections and transfers.

In accordance with a resolution of Council, and in the absence of any objections, the election and transfer of the candidates to the class indicated will be confirmed fourteen days after the date of circulation of this list. Any objections or communications concerning these elections should be addressed to the Secretary for submission to the Council.

Direct Election to Member

KNOWLES, Lieutenant-Colonel Royston, R.E.M.E. *Malvern Wells, Worcs.*

Transfer from Associate Member to Member

CHALMERS, David. *Wirral, Cheshire.*
ISHERWOOD, Charles Fitton, B.Sc. *Bogota, Colombia.*
POWELL, Claud. *London, S.W.7.*
SILBERMAN, Herbert. *Johannesburg, South Africa.*
SIMPSON. David, Dip. Eng. *Kingskettle, Fife.*
SMITH, George Donald, C.G.I.A. *Northampton.*

Direct Election to Associate Member

COTCHER, Alfred L. *London, S.W.3.*
DULLFY, Peter John. *Ingatestone, Essex.*
FINCH, Cyril *London, E.18.*
GUMERY, Edward Alfred. *Watford, Hertfordshire.*
KEELING, Alfred Maxwell. *Bath, Somerset.*
LAISHLEY, Frederick Herbert. *Solihull, Warwickshire.*
LONGDEN, Thomas Edwin. *Wickford, Essex.*
MEACHEM, Wing-Commander, William George, R.A.F. *Leighton Buzzard.*
PARFREY, William Michael John. *Stevenage, Hertfordshire.*
PHIPPARD, Major Roy Gordon, R. Sigs. *Catterick Camp, Yorkshire.*
SCHOFIELD, Squadron Leader Arthur Ernest, R.A.F. *Bedford.*
SMURTHWAITE, Squadron Leader Richard, R.A.F. *Middlesbrough.*
STEWART, Major Peter Hartley, R.A. *Shipton Bellinger, Hampshire.*
TAYLOR, Roy Edmund. *Windsor, Berkshire.*
WITHAMS. Kenia Frederick, B.Sc. (Eng.). *Chatham, Kent.*

Transfer from Associate to Associate Member

HOWATT, Lieutenant Commander George D., R.N. *Fareham, Hampshire.*

Transfer from Graduate to Associate Member

ABHYANKAR, Flight Lieutenant Moreshwar K., I.A.F. *Poona, India.*
AVINOR, Michael, Ph.D. *Haifa, Israel.*
BEX, John Charles Arnold, B.Sc. *London, S.W.4.*
CROUCH, Malcolm Jasper. *Southampton, Hants.*
DALTON, Walter James. *London, W.4.*
FLINT, Reginald Arthur. *Leicester.*
HARCHARAN SINGH. *Harlow, Essex.*
HOWARTH, Edwin. *Torpoint, Cornwall.*
JAMIESON, Clifford Park. *Liverpool.*
LLOYD, Gwilym Bevers, B.Sc. *Bath, Somerset.*
MAH, Seck Wah. *Kuala Lumpur, Malaya.*
NORTHMORE, Squadron Leader William, R.A.F. *Basingstoke, Hampshire.*
PASFIELD, Arthur Edmund. *Great Malvern, Worcestershire.*
REYNOLDS, John Frederick, B.A. (Cantab.). *Tewkesbury, Gloucestershire.*
ROGERS, David Walter William. *Weymouth, Dorset.*
SCRUSE, Stanley Warren. *London, N.13.*
THOMAS, Philip Robinson. *Bedford.*
WOLFE, Brian Sinclair. *London, N.2.*
ZAHEDY, Javad. *Weybridge, Surrey.*

Direct Election to Associate

BIRD, Stanley Granger, B.Sc. *London, E.17.*
GOODMAN, Robert Neville. *London, N.W.4.*
KIRK, Ivor Osborne. *Aylesbury, Buckinghamshire.*
WATERS, Ian Morley. *Ely, Cambridgeshire.*

Direct Election to Graduate

ALBERTELLA, Victor John. *South Harrow, Middlesex.*
ARTHUR, Geoffrey. *Croydon, Surrey.*
CARPENTER, John Radcliffe. *Chelmsford, Essex.*
COOPER, Cedric. *Birkenhead.*
CROMPTON, Craig Pickop. *Stroud, Gloucestershire.*
GASKING, John David. *Upminster, Essex.*
HUGHES, Owen Cecil. *Beckenham, Kent.*
JANES, Roger Heslam. *Cheltenham, Gloucestershire.*
MACHU, Gerald James. *Guildford, Surrey.*
NIGHTINGALE, Clive Richard. *Bromley, Kent.*
OLSZEWSKI, Peter Hans. *Naenae, New Zealand.*
PETTIT. Christopher Richard. *West Kirby, Cheshire.*
PHILLIPS, Robert Wilfrid. *Waltham Abbey, Essex.*
POND, Barry John. *Tadley, Hampshire.*
READ, Brian John. *Luton, Bedfordshire.*
SASTRY, Vendantum Lakshmipathy, B.E. *Bangalore, India.*
SATTERTHWAITE, Edward. *Coventry, Warwickshire.*
STEVENS, John Francis. *Emsworth, Hampshire.*
THODAY, Robert David Cyril. *Reigate, Surrey.*
THOMPSON, Henry Victor. *Bletchley, Buckinghamshire.*
THOMPSON, Ronald Leonard. *Southampton, Hampshire.*
TWIST, Barry Hilton. *London, S.E.26.*
WARBURTON, Harold. *Basingstoke, Hampshire.*

Transfer from Student to Graduate

ARNETT, Leslie Frederick. *Great Malvern, Worcestershire.*
DIXON, Cyril Ernest. *Luton, Bedfordshire.*
HENDY, Jeffrey Alan James. *Southampton, Hampshire.*
LEWIS, Arthur Dennis. *Berbera, N.T., Somali Republic.*
RICHARDSON, Kenneth Charles Smart. *Kilbarchan, Renfrewshire.*
ROMAINE, David Albert Frederick. *Bracknell, Berkshire.*
SHARMAN, Harold. *Currie, Midlothian.*
TAYLOR, Edwin Leslie. *Marlow, Buckinghamshire.*
TSANG, Kwan Cheuk. *London, N.2.*
VOJDANI, Mostafa. *Abadan, Iran.*
VOVIDES, Andreas Christou. *London, N.8.*
WILLIS, Roy Frederick. *Totton, Hampshire.*
YASHWANT DEVA, Captain. *Ludhiana, India.*

STUDENTSHIP REGISTRATIONS

The following students were registered at the March and May meetings of the Committee. The name of a further 25 students registered at the May meeting will be published later.

BASSIL, Geoffrey Charles. *St. Albans.*
CLARK, Thomas. *Epping, Essex.*
JACKSON, Frank Michael. *London, W.C.1.*
KENWARD, Michael. *Wallington, Surrey.*
KILLEEN, Christopher T. *Dublin.*
KING, Albert Edward. *Southampton.*
LEES, William, M.A., *Axbridge, Somerset.*
McNAMEE, Brian Michael. *Cabra, Dublin.*
MANNING, W. A. *Auckland, New Zealand.*
MANNION, Martin, B.Sc. *Dublin.*
MILLS, Ivan R. *Clacton-on-Sea.*
MITHANI, Kishore S. *London, N.13.*
MORGAN, Brian Wilson. *Cardiff.*
NJOKU, Emmanuel A. C. *Lagos.*
O'DONOGHUE, Thomas K. *London, N.2.*
PATON, Alan J. *Ffestiniog, Merioneth.*
PEARSON, Nicholas David. *Stafford.*
POLLING, Harvey. *London, N.3.*
POORL, Norris G., B.E.M. *Weston-super-Mare Somerset*
RAMACHANDRAN, Chinniah. *Colombo.*
RANASINGHE, Herbert G. S. *Colombo, Ceylon.*
RATHI. Devdas N., B.E., M.E. *Nagpur, India.*
REYNOLDS, Lyn S. *Par, Cornwall.*
RICHARDS, Harold. *Newport, Monmouthshire.*
ROBERTS, William Marshall. *Edinburgh.*
SCOTT, Peter. *Barnet, Hertfordshire.*
SELBY, John H. R. *Weston Favell, Northants.*
SETHURAMANAIAH, H., B.Sc. *Deccan, India.*
SINGER, Lester Simeon. *Auckland.*
STEPHENSON, John. *London, E.4.*
SUDUL, Keith M. *Feltham, Middlesex.*
TAYLOR, John Chisholm. *South Shields.*
THAKE, Phillip Edward. *Leicester.*
TOOZE, Michael John. *London, S.W.2.*
UPADHYAY, N., B.Sc. *Faizabad, India.*
VAID, Joginder Nath. *London, N.W.11.*
VENKATACHALAM, G., B.Sc. *East Punjab.*
*VINNER, Reuben. *Tel Aviv, Israel.*
WESTON, Reginald I. *St. Osyth, Essex.*
WOO SWEE SENG. *Kuala Lumpur.*
WOOD, Walter Alfred. *Abingdon, Berks.*
WRIGHT, Douglas Irvine A. *Coventry.*
YEANG KIAT WONG, *London, N.W.11.*

ADEOSUN, Stephen Eliot. *Lagos.*
ASADULLAH. S. *Karachi, Pakistan.*
BABU RAO SHENOY, S., B.Sc. *Mysore State, South India.*
BARMAN, Shakti Prasad, B.Sc. *Calcutta.*
BAYFIELD, Ronald. *Brighton, Sussex.*
BRIERLEY, Norman Leslie. *London, S.E.19.*
CERESOLE, P. L., B.Sc. (Eng.). *London, S.W.5.*
CURTIS, Alan Paul. *London, W.13.*
DAGLISH, Anthony. *Seascale, Cumberland.*
DATE, Prabhakar Canesh, B.Sc.. *Bombay.*
de RYKE, Cornelis. *Kilkenny, South Australia.*
DESAI, Shailesh T., B.E. *Fleet, Hampshire.*
DILLICAR. Colin R. *Auckland, New Zealand.*
DOYLE, Thomas E. M. *Bracknell, Berkshire.*
ENDERBY, David John. *Lincoln.*
GANGANNA, Honnappana Hall Veeranna, B.Sc., M.Sc. *Mysore State, India.*
GUILDFORD, Leslie. *Burgess Hill, Sussex.*
GUPTA, Ramesh C., B.Sc. *Hoshiarpur, India.*
HADJIVASSILIOU, Charalambos. *London, N.4.*
HARI SHANKAR, B.Sc., M.Sc. *Delhi.*
HARIHARAN, T., M.Sc., B.Sc. *Cochin, India.*
HORTON, Laurence, B.Sc. *London, S.W.2.*
JENKINS, Kingsley. *Colchester, Essex.*
JINNAH, Karim D., B.Sc. *Poona, India.*
JOGLEKAR, Sudhakar Bhikaji, B.Sc. *Bombay.*
JONES, Alec. *Penarth, Glamorgan.*
JOSEPH, P. L., B.E. *Ernakulam, South India.*
KATENDE, Francis Xavier. *Chelmsford.*
KING Robert Charles. *Cirencester.*
KOOSHKABADI, Gharab. *Abadan, Iran.*
KULKARNI, Govind Dattatraya. *Poona.*
KUSHWAHA, Ramesh. *Kanpur, India.*
LEES, Frank P. *Salisbury, Southern Rhodesia.*
LIDBETTER, Frank E. *Welling, Kent.*
LUKTUKE, Raviprakash D., B.Sc. *Poona.*
McDERMOTT, Patrick A. *Cottingham, Yorks.*
McELROY, Antony. *Greenock, Renfrewshire.*

* Reinstatement.

U.D.C. 681.142

Data Collection and Distribution

By
D. J. DACE †

Presented at the Symposium on Data Transmission in London on 3rd January 1962.

Summary: The paper outlines the evolution of data handling and processing methods employed by a group of insurance companies in day-to-day operation, culminating in the adoption of a fully integrated computer system. Field trials of a telex link using live data and subsequent extension of these trials to embrace the use of high-speed transmission equipment are described. A summary of the transmission errors encountered during the later trials is given together with the approximate comparative costs of a telex and high-speed system.

1. Introduction

Prior to 1953, the author's company had relied largely on manual clerical methods, each branch office being primarily responsible for the issue of new policies and the production of renewal notices and agents' accounts, although for some time fire renewal notices had been prepared using an address plate. The production of up-to-date statistics for management, however, was largely impossible due to the time-consuming task of collating and analysing the information provided by Branches.

It had been apparent for some time that there was a growing need for a more economical and more efficient system and, in 1953, after many months of planning and investigation the first steps were taken towards an integrated system which was to be, at that time, probably the most comprehensive in Great Britain and, possibly, in the world.

Detailed study had shown that at that stage, electronic computers were still comparatively untried for such large scale commercial data processing. The initial step, therefore, was to establish centralized punched-card installations at Croydon, Exeter and Rickmansworth, the last named being responsible for the additional tasks of collating statistics from all three units and producing reports for management, issuing dividend warrants and handling staff salaries for the entire group of companies. These punched card installations were to bridge the gap until such time as computing equipment of sufficient power and reliability became available and the card files could then, by their very nature, simplify the problem of transcribing the Company's records to magnetic tape.

The preparation of punched cards for an organization comprising many branch offices naturally lends itself to the establishment of centralized punching and verification of data on economic grounds. The source documents were sent to the central units from the branches daily and the tabulated results of the processing were returned to the branch concerned.

The first method of data transmission was therefore by post. It was cheap and fairly reliable; the only transmission errors occurred when a branch forgot to reverse the address card in the transparent aperture of the mailing bag and promptly received their mail back again the following morning.

2. The Need for Data Transmission

The punched card system worked, and is still working, very efficiently but with such a system there is inevitably a delay in getting source data from branches, punching, verifying, printing and sending the document back to the branch. Looking ahead to the computer it was realized the immense speed of the machine would be wasted unless the time needed to receive, punch and verify the data could be reduced. All stages which the data must go through before reaching a form acceptable to the computer must be streamlined.

It was realized the only way to achieve this result and so provide an efficient service to customer and management was to decentralize data preparation, that is have the data translated, at the branch, as it became available, into a medium which could be fed directly into the computer system and to ensure that it reached the computer centre in the minimum of time. The answers to the last problem seemed to be data transmission.

3. Experiments with Telex

Early in 1959 the G.P.O. was approached on the question of data transmission. At this time telex was the only facility available and so this was adopted to give practical experience in operation procedures and reveal the type of errors that this form of data transmission would produce.

† Commercial Union Assurance Company Ltd., Mechanization Department, Exeter, Devon.

The eventual result was to establish a link between the London City Office and the punched card unit at Croydon. Tape was prepared on the teleprinter keyboard by the branch, verified by a call-over of the hard copy produced against the original source document, and then transmitted to Croydon. At the unit the tape was fed directly into an I.C.T. 1036 tape-to-card converter, the necessary cards being thus automatically produced.

Results achieved by this experimental link were very encouraging and showed that, over short distances, the use of telex was practical.

In the spring of 1960 the Company, after a year's study of existing and proposed computers, ordered an English Electric KDP 10. It was decided that this would be housed in a building to be specially designed and built in Exeter.

The use of Devon as a location meant long distances to transmit data and therefore to economize on the cost of "Tariff H" private telegraph circuits a system was drawn up whereby branches would be grouped in geographical zones with a main transmitting branch and the other branches acting as satellites. Each satellite would have a private line to the main branch and each main branch a private line to the computer centre at Exeter. The satellites would transmit data to the main branch who would then re-transmit to Exeter.

Tariff H private telegraph circuits were to be used in the absence of experience of normal telex errors over long distances.

To date the experiences with telex had been most satisfactory. The cards produced at Croydon were accurate and error free. The G.P.O. had provided excellent servicing facilities on the very few occasions we had had need to call upon them.

4. Consideration of High-speed Transmission

There were however many disadvantages. The KDP 10 computer was designed to work with a seven-unit code and thus all tape would need to be converted either at input or by a subsequent conversion programme. Speed of transmission was low, only 50 bauds, which would mean a total theoretical daily transmission time of some 60 hours for our needs. There was no way of checking how accurate the information was, since telex code has no parity and the transmission system no provision for correcting or even for detecting errors.

By this time several manufacturers were working on high speed transmission systems and a survey was taken of all equipments which were available or likely to be available by the time the computer was delivered in January 1962.

The requirements such a system would have to fulfil were:

(a) Be capable of transmitting at a speed greater than 350 bauds in order to reduce call-times.

(b) Be capable of handling 1 in. wide tape utilizing a 7-unit code.

(c) Have an undetected error rate of not worse than 1 in 10^6 characters.

(d) That the system complete with necessary tape preparation equipment be of comparable cost with a telex network as envisaged earlier.

Arrangements were subsequently made with two companies for the loan of suitable equipment for field trials.

5. Cost

One of the conditions of accepting high-speed transmission equipment was that the cost of purchasing transmission and data preparation equipment plus the cost of calls and maintenance would not exceed the annual rental of telex equipment and private wire telegraph circuits. A five-year period was allowed over which to spread the capital cost of equipment.

The following costing given is based upon 43 branches transmitting 1½ million characters per day and utilizing 80 data preparation machines (Table 1).

(a) For telex:

(i) Rental of a country wide network of "Tariff H" private telegraph circuits (Fig. 1).

(ii) The rental cost of the necessary telex equipment on which to prepare the data.

Table 1

Comparison of cost of telex "Tariff H" private wire network and high-speed transmission using switched telephone network

	Telex	High-speed data transmission
Cost of calls	—	£5000
Private wire rental	£21 000	£2000
Equipment hire	£19 500	—
One-fifth purchase price of transmission and preparation equipment	—	£24 000
Maintenance of transmission and preparation equipment	—	£7000
	£40 500	£38 000

(b) For high-speed equipment:

(i) Capital cost of data preparation equipment.

(ii) Cost of data transmission equipment—capital cost.

(iii) Maintenance on preparation equipment—per annum.

(iv) Maintenance on transmission equipment—per annum.

(v) Cost of calls on public network. This has been quoted on normal rates. Subscriber trunk dialling will save approximately £1000 per annum. Speed of transmission has been taken as an effective 350 bauds.

(vi) Cost of hire of a full time private line between London and Exeter for the only two-way link in the system.

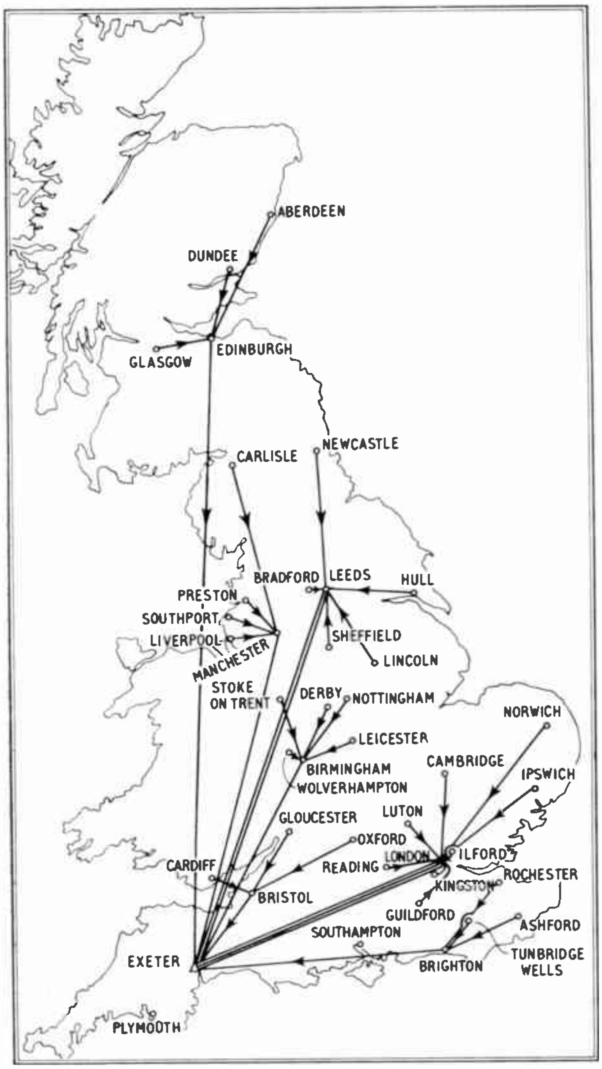


Fig. 1. Proposed "Tariff H" private telegraph wire system.

6. Data Preparation

To prepare data at branches and use high speed transmission it was necessary to equip the branches with a machine capable of producing a media which could be transmitted, since the equipment used in the field trials cannot incorporate a keyboard as does telex. The need was for a machine which would prepare the transmission media as a by-product of normal document preparation at the branch, and which could easily be verified for accuracy by a "call-over" against the source information.

The answer was to use a typewriter which would produce either a card or a punched paper tape. Paper tape was chosen for the following reasons.

(a) Speed of input of paper tape to the computer is twice as fast as card input.

(b) Both preparation and transmission equipment were cheaper for paper tape.

(c) The advantage of using cards, i.e. their ease of sorting prior to input, was not needed.

(d) Paper tape was itself much cheaper than punched cards.

In this way the old punched card drawback of needing three different sets of key taps for any piece of information fed into the system (original typing, punching, verifying) was superseded by a system needing only one set.

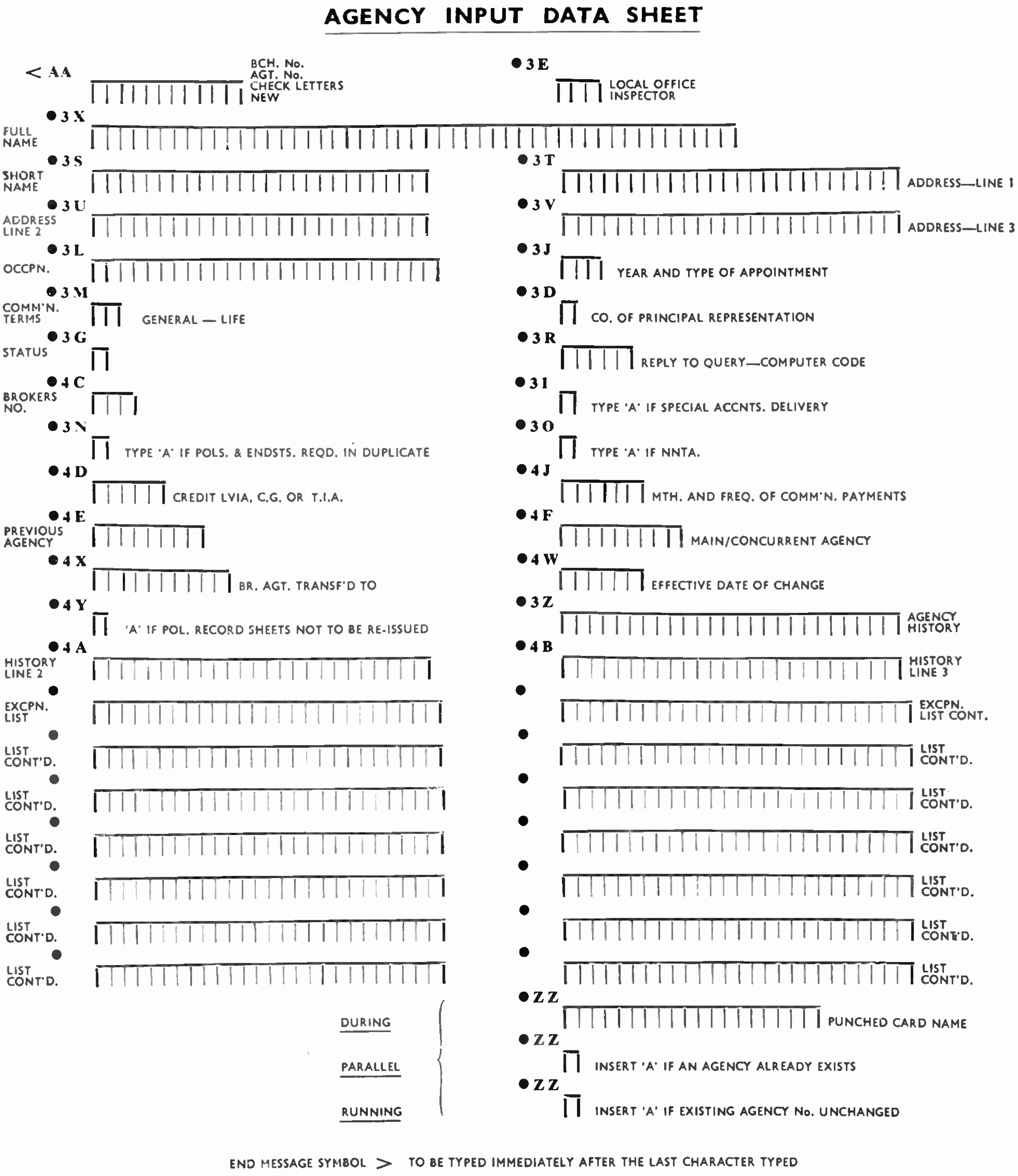
A survey was taken of all known data preparation equipment and two companies supplied suitable equipment on a free trial basis.

Special forms have been designed for use by Branch offices in data recording and punching, the form relating to new agency being shown in Fig. 2. The individual "boxes" are designed to contain the maximum permissible item size and the form is preprinted with the SM, EM and ISS control symbols and identifying line labels as an aid to the typist.

7. Field Trials with High-speed Transmission

By early 1961 data preparation equipment was installed and working and a set of high-speed transmission equipment available. The only hold up was G.P.O. approval to use the public network. Arrangements were made to start the trials on a part-time private wire but before the line was available limited approval was given for a test over the public network.

The equipment used for the first field trials was a version of the A.T. & E. "Swift" data transmission system. Input of data to the transmitter is by means of a Ferranti 100 character per second paper tape reader, output from the receiver being recorded by a Teletype BRPE 110 character per second punch. The basic transmission speed of this phase-modulated system is 750 bauds but, due to the transmission of extra parity bits for error detection and correction,



AGENCY INPUT DATA SHEET

< AA — BCH. No. AGT. No. CHECK LETTERS NEW

● 3 E — LOCAL OFFICE INSPECTOR

● 3 X — FULL NAME

● 3 S — SHORT NAME

● 3 T — ADDRESS—LINE 1

● 3 U — ADDRESS LINE 2

● 3 V — ADDRESS—LINE 3

● 3 L — OCCPN.

● 3 J — YEAR AND TYPE OF APPOINTMENT

● 3 M — COMM'N. TERMS — GENERAL — LIFE

● 3 D — CO. OF PRINCIPAL REPRESENTATION

● 3 G — STATUS

● 3 R — REPLY TO QUERY—COMPUTER CODE

● 4 C — BROKERS NO.

● 3 I — TYPE 'A' IF SPECIAL ACCNTS. DELIVERY

● 3 N — TYPE 'A' IF POLS. & ENDSTS. REQD. IN DUPLICATE

● 3 O — TYPE 'A' IF NNTA.

● 4 D — CREDIT LVIA, C.G. OR T.I.A.

● 4 J — MTH. AND FREQ. OF COMM'N. PAYMENTS

● 4 E — PREVIOUS AGENCY

● 4 F — MAIN/CONCURRENT AGENCY

● 4 X — BR. AGT. TRANSF'D TO

● 4 W — EFFECTIVE DATE OF CHANGE

● 4 Y — 'A' IF POL. RECORD SHEETS NOT TO BE RE-ISSUED

● 3 Z — AGENCY HISTORY

● 4 A — HISTORY LINE 2

● 4 B — HISTORY LINE 3

● — EXCPN. LIST

● — EXCPN. LIST CONT.

● — LIST CONT'D.

● — LIST CONT'D.

● — LIST CONT'D.

● — LIST CONT'D.

● — LIST CONT'D.

● — LIST CONT'D.

● — LIST CONT'D.

● — LIST CONT'D.

● — LIST CONT'D.

● — LIST CONT'D.

● — LIST CONT'D.

● — LIST CONT'D.

● Z Z — PUNCHED CARD NAME

● Z Z — INSERT 'A' IF AN AGENCY ALREADY EXISTS

● Z Z — INSERT 'A' IF EXISTING AGENCY No. UNCHANGED

DURING
PARALLEL
RUNNING

END MESSAGE SYMBOL > TO BE TYPED IMMEDIATELY AFTER THE LAST CHARACTER TYPED

Fig. 2. Specimen branch input sheet for new agency data.

the net information transfer rate is reduced to 437 bauds or 62·5 KDP 10 characters per second.

The arrangement of the information and parity bits is set out in Table 2; the five parity bits are of course generated by the transmitter circuitry and do not appear in the paper tape.

The data transmitted were of two sorts:

(i) Actual data relating to new policies and cash payments.

(ii) Test data contained on loops which were used to fill in the time when no actual data were available.

Table 2

Information and parity bits

1 0 0 0 0 1 0 = KDP 10 Code Letter "B"

	I_1	I_2	I_3	I_4	I_5	I_6	I_7	P_1	P_2	P_3	P_4	P_5	
Parity Check 1	×			×	×	×		×					Parity Bit $P_1 = 0$
,, ,, 2	×	×			×		×		×				,, ,, $P_2 = 1$
,, ,, 3	×	×	×			×				×			,, ,, $P_3 = 0$
,, ,, 4			×	×	×		×				×		,, ,, $P_4 = 0$
,, ,, 5		×	×	×		×	×					×	,, ,, $P_5 = 1$

With this type of code all single bit errors in any character can be automatically corrected as also can some multiple-bit errors. In all, some 90% of detected errors can be corrected in this manner, the remainder causing the automatic stopping of both receiver and transmitter. This is achieved quite simply by cutting off the ARQ continuous tone signal which, during normal transmission is sent back down the line from receiver to transmitter using the lower part of the frequency band. Removal of this signal causes the transmitter to stop and is used also in the following circumstances:

(i) Low level of paper tape at receiver.

(ii) Use of "emergency stop" button at either terminal.

(iii) Power or equipment failure.

(iv) Line disconnection.

When transmission has been interrupted, the transmitter operator merely back spaces the tape to the previous inter-message gap prior to re-starting transmission.

The tests were carried out between Brighton and Exeter, transmission being in one direction only. Calls were normally established from the receiving end at Exeter using the public telephone network but on a P.U.T. basis (prolonged-uninterrupted), thus overcoming the problem of the "three minute" pips.

The equipment was operated by the company's own staff from the typing pool at Brighton and the clerical section of the punched card unit at Exeter. None had had any previous experience of operating paper tape equipment.

The test loops contained the following information:

Test tape A: Alternate holes.

B: Random numerals and alphabetical characters.

C: Holes 1 and 7 only punched.

D: All holes punched.

E: Alternate seven holes punched and seven blanks.

F: Repetitive blocks.

G: Blank tape.

The actual data were used in the punched card unit and the transmission has now completely replaced post as a method of sending all new policy and cash data between unit and branch.

Figures of undetected errors were arrived at by checking every character punched. No reliance was placed on counters to detect errors and in this way it was possible to be certain that any error occurring throughout the entire system, from data preparation at the branch to the data being punched at the unit, was detected.

The checking took the form of a visual comparison between the transmitted and punched tapes. The two tapes were of different colours, yellow and red respectively, and so errors were easily detectable.

Transmissions on which the results given in Table 3 are based were carried out four times a day at the approximate times shown. The duration of each test was approximately 15 minutes. Operators practised tape changes whenever a spool was low.

Results were kept of number of characters transmitted, number of detected uncorrected errors, number of undetected errors and totals of detected corrected errors.

8. Error Detection and Correction

Errors are inevitable in any system in which the human element enters and they also occur occasionally in systems where it is not. Provision must of course be made for the detection and correction of these errors, preferably before the data reach the computer. These provisions must be as foolproof as possible but without resorting to too complex a system; complicated correction procedures, especially for typists or punch operators, often only serve to engender further errors of their own.

Errors arising during data preparation fall into three main categories:

(a) those noticed by the typist immediately;

(b) those detected after having typed the incorrect item, and possibly several other items, but without having yet completed the entire message;

(c) those found subsequently during checking.

The first of these variations is dealt with by erasing the incorrect character on tape, the depression of the "delete" key simultaneously causing a box to be printed around the relevant printed character, and then re-typing the data.

In the case of the second form the message is immediately terminated by the error code and the EM symbol. On input of paper tape to the computer, the program automatically checks the last character before the EM symbol and rejects the message if it discovers the error code.

The complete message is then retyped.

Where the third type of error is concerned the complete message is retyped with the addition of a "correction" character. This allows the computer to ignore the earlier message. Any number of corrections can be catered for by this method.

Transmission errors similarly fall into three distinct divisions:

(a) simple errors, usually affecting only one bit per character;

(b) more complex interference, including line interruptions;

(c) the fortunately rare errors that go undetected.

Those in category (a) are detected and automatically corrected in the receiving terminal before being punched into paper tape. Where the errors are of such complexity or severity, as in type (b), that they are beyond the scope of the parity checking system, transmission is automatically halted. The transmitter operator then retransmits the message in which the break occurred. The computer program will detect the redundant partial message structure in the KDP 10 computer system in which the group of items comprising a message are preceded by an SM (start message symbol) and followed by an EM (end message symbol). The redundant partial message is detected by the computer program due to the absence of the EM symbol.

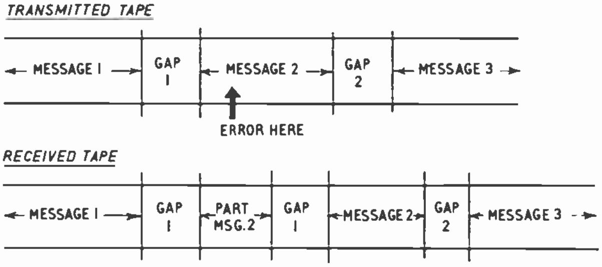


Fig. 3. Illustrating detection and correction of errors.

An example is given in Fig. 3.

A similar method is used when changing a tape during a transmission at the receiving end. The receiving operator stops her machine when her tape is getting low. This interrupts the ARQ signal and is detected at the transmitter, the operator of which, then backs the tape and re-transmits from the end of the last good message. This procedure ensures that a message is not lost on tape changes at the receiver.

9. High-speed Transmission Trials

The results shown in Table 2 are for the period 27th July to 8th December 1961. Some days results have been excluded as follows:

(a) The period 1st–7th September had some 24 undetected errors in 650 000 characters, but not one detected uncorrected error. These results were due to a transistor failure which effectively did away with the error detection facility. Even so the overall error rate was $3{\cdot}7 : 10^5$.

(b) The second period omitted was when the regular operator was sick. The replacement girl tried splicing tapes, with apparently little success, more glue depositing itself on the tape reader than the tape. 78 errors were found in some 1½ million characters. As soon as the regular operator returned the error rate returned to its normal level.

(c) Two further transmissions are not included. These were when line conditions were so bad that one operator closed the call down, and retransmitted at a later period. Ten undetected errors were discovered in 3 transmissions.

Table 3
Transmission trial results 27th July–8th December 1961

	All data				Test tape				Cash/new policies			
Time of transmission	Number of characters	Detected/ uncorrected errors	Undetected errors	Undetected error ratio	Number of characters	Detected/ uncorrected errors	Undetected errors	Undetected error ratio	Number of characters	Detected/ uncorrected errors	Undetected errors	Undetected error ratio
9.30 a.m.	3 382 000	148	8	2·36 : 10^6	3 358 000	148	8	2·38 : 10^6	24 000	0	0	∞
12.00 noon	2 960 000	106	1	0·33 : 10^6	2 744 000	85	0	∝	216 000	21	1	4·62 : 10^6
2.15 p.m.	3 397 000	110	20	5·88 : 10^6	3 154 000	100	19	6·02 : 10^6	243 000	10	1	4·11 : 10
3.45 p.m.	2 796 000	53	4	1·43 : 10^6	2 239 000	48	4	1·78 : 10^6	557 000	5	0	∞
Overall	12 535 000	417	33	2·63 : 10^6	11 495 000	381	31	2·69 : 10^6	1 040 000	36	2	1·92 : 10

Even if all the above errors were included, the final result would show an error rate of 1 : 10^5.

It is interesting to note the consistency of the error rate throughout the trial period:

27th July–31st August— 4 000 000 character, error rate 3 : 10^6

8th September–19th October— 4 800 000 character, error rate 3·3 : 10^6

Overall 12 500 000 character, error rate 2·63 : 10^6

Error rates are expressed in characters. The corresponding bit error rate is twelve times as good.

In 8 500 000 characters we have had 2334 errors automatically corrected, 364 detected but uncorrected and 21 undetected. This shows that:

85·84% of errors have been automatically corrected,

13·39% of errors have been detected,

0·77% of errors have been undetected.

Put a different way it shows that on average;

undetected errors have occurred every 2 hours,

detected errors every 8 minutes,

and corrected errors about every minute.

It must be stressed that in view of the experimental nature of the exercise many calls were made over extremely bad lines, when in normal operating circumstances the call would have been broken down and set up again. In one such call 42 errors were detected and 174 automatically corrected, although no errors remained undetected.

Of the 33 errors in transmission, 22, i.e. 66%, would have failed the parity check on the KDP 10 computer, thus giving an undetected error rate over the system of less than 1 in 10^6.

10. Independent Tests

On the 14th September the G.P.O. carried out an independent series of tests over various switched networks.

The normal maximum number of exchanges through which any call from one of our Branch Offices to Exeter will go is 3. As can be seen from the results given in Table 4, on most calls as many as five exchanges were involved. Tests 1 to 8 produced results which corresponded closely to those produced by our main series of tests.

The A.T.E. ARQ tone alternates between 340 c/s and 410 c/s at 1 second intervals. If this tone is cut off the transmission ceases and the "no signal" indication is given at the receiver.

In tests 9–13 the attenuation at the lower frequency end was very high. G.P.O. have recorded it as "more than 30 dB". The result was that at 340 c/s the ARQ was being cut off but at 410 c/s the receiver managed to pick it up satisfactorily, hence the "no signal" indication at 1 second intervals. The receiver sensitivity on being increased (by manual tuning) was found to be perfectly capable of operation at this attenuation.

11. Conclusion

Investigations and trials of high speed data transmission systems coupled with decentralized data preparation have demonstrated the practicability of such a system for every day use. It is true that the normal postal service is cheap and efficient but the cost of a high-speed line system is regarded as justified by the better service that can consequently be offered to the customer and to the management.

Plans are already being formulated for the stage after all home offices in England, Scotland, Ireland

Table 4

Transmission results with experimental routings, Brighton to Exeter, conducted by G.P.O.

Routing	Speech Conditions	Direction	Mid-frequency band attenuation (1000 to 1600 c/s) Approx. loss—dB	Number of characters transmitted	Errors		
					Detected and Corrected	Detected and not Corrected	Un-detected
1. Normal	Good	EX—BR	13·5	56 000	3	Nil	Nil
		BR—EX	12·0				
2. Normal	Good	EX—BR	13·0	56 000	Nil	Nil	Nil
		BR—EX	10·5				
3. Exeter—London—Brighton	Good	EX—BR	14·1	53 000	4	Nil	Nil
		BR—EX	12·0				
4. Exeter—Bristol—London—Brighton	Good	EX—BR	20·0	90 000	31	8	Nil
		BR—EX	17·5				
5. Exeter—Birmingham (Newhall)—Birmingham (TK)—London—Brighton	Good	EX—BR	17·5	64 000	17	1	1
		BR—EX	18·7				
6. Exeter—Bristol—Birmingham—London—Brighton	Good	EX—BR	13·7	56 000	29	9	Nil
		BR—EX	19·7				
7. Exeter—London—Horsham—Brighton	Very Good	EX—BR	19·2	56 000	Nil	Nil	1
		BR—EX	15·0				
8. Normal	Good	EX—BR	11·6	19 000	13	Nil	Nil
		BR—EX	10·5				
Summary of errors				450 000	97	18	2
					(error rate 4·4 : 10⁶)		
9. Exeter—Bristol—Leicester—London—Brighton	Fair improved to Good	EX—BR	19·3	51 000	31	3 + 63 No Signal	3
		BR—EX	18·5				
10. Exeter—Bristol—London—Horsham—Brighton 1st Call	Good	EX—BR	23·5	"No Signal" at approx. 1 second intervals See text			
		BR—EX	21·5				
11. As No. 10, 2nd Call	Good	EX—BR	20·7				
		BR—EX	25·0				
12. Exeter—Bristol—London—Portsmouth—Brighton	Good	EX—BR	22·5				
		BR—EX	19·5				
13. Exeter—Bristol—Liverpool—London—Brighton	Good	EX—BR	Not taken				
		BR—EX	Not taken				

and Wales have been linked to the computer centre at Exeter. These plans include the provision of links with those overseas branches whose volume of business does not warrant the installation of separate computer systems.

12. Acknowledgments

The author is indebted to Mr. F. C. Knight of Commercial Union for permission to publish this paper. He also acknowledges the assistance provided by Mr. I. A. Edmonds, of the English Electric Co.'s Data Processing Division both in its preparation and in the early investigations of available data transmission and data preparation equipments.

Manuscript received by the Institution on 25th September 1961. (Paper No. 725.)

U.D.C. 621.317.72 : 621.372.2

Reflection-Coefficient Curves of Compensated Discontinuities on Coaxial Lines and the Determination of the Optimum Dimensions—Part 2

By

A. KRAUS, DR. ING.†

Summary: It is shown that the field inhomogeneities caused by supports are limited to the border region. If the compensation is made in this zone the frequency-dependent reflection is very low. The optimum dimensions for supports compensated in their border regions are derived from a series of measurements.

1. Introduction

In an earlier paper,[1] curves of reflection coefficients and optimum dimensions of dielectric supports in coaxial lines are given, the optimum characteristics being obtained by varying the depth of penetration of the dielectric into the conductors. These supports have a larger diameter ratio than a homogeneous line of equal characteristic impedance, because higher inductance is necessary to compensate for the spurious components resulting from the inhomogeneous fields. It is typical that the diameter ratio D/d of these supports is always

$$\frac{D}{d} > \exp\left(\sqrt{\varepsilon}\,\frac{Z}{60}\right) \qquad \ldots\ldots(1)$$

where Z is the characteristic impedance of the adjoining line (Fig. 1). Even with optimum dimensions the supports present relatively large reflection. This is explained by the detail structure, i.e. the configuration of the equivalent four-terminal networks of the support, which is compensated only on the average, while remaining local discontinuities cause residual reflection.

In order to reduce the reflection further, the supports must be compensated not only on the average, but in individual parts, so that every spurious capacitance is locally compensated by a suitable inductance. With adequate subdivision into compensated sections, a support presenting still smaller reflection would be possible. However, the local distribution of the spurious capacitances must be determined to make this compensation possible.

2. Supports with Locatable Spurious Capacitance

The spurious capacitance of a line of characteristic impedance Z changing directly from the dielectric ε_0 to a dielectric ε_1 lies in the plane of transition, where the diameters of the conductors must change suddenly in order that the characteristic impedance Z remains constant (Fig. 2). The dimension of the inhomogeneous fields in the direction of the line axis is relatively small. An equivalent lumped spurious capacitance C_s may be calculated.[2] When determining the optimum compensation by measurement, it is good practice to measure the reflection as a function of frequency. A discrepancy between calculated and measured values exists at high frequencies, because the approximation of the distributed inhomogeneous field by a constant lumped capacitance is fairly inaccurate. If the breadth of the support in the test item is made approximately equal to the diameter of the outer conductor ($B \simeq D$), then the spurious capacitances are spaced apart about $D\sqrt{\varepsilon_1}$. This distance can be assumed to be large enough to exclude interaction of the fields. If the calculated characteristic impedance in the support is made equal to the characteristic impedance Z of the adjoining homogeneous line, i.e.

$$\frac{D}{d} = \exp\left(\sqrt{\varepsilon}\,\frac{Z}{60}\right) \qquad \ldots\ldots(2)$$

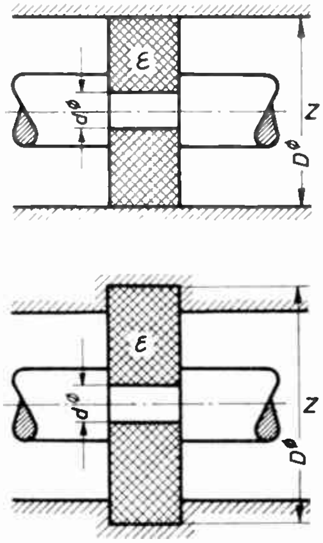


Fig. 1. Dielectric supports the diameter ratios of which are enlarged for compensation:

$$\frac{D}{d} > \exp\left(\sqrt{\varepsilon}\,\frac{Z}{60}\right)$$

† Rohde & Schwarz, Munich.

the equivalent circuit of the support may be represented by a homogeneous line of characteristic impedance Z which is loaded by two spurious capacitances C_s spaced apart about $B\sqrt{\varepsilon_1}$ (Fig. 2 (*b*)).

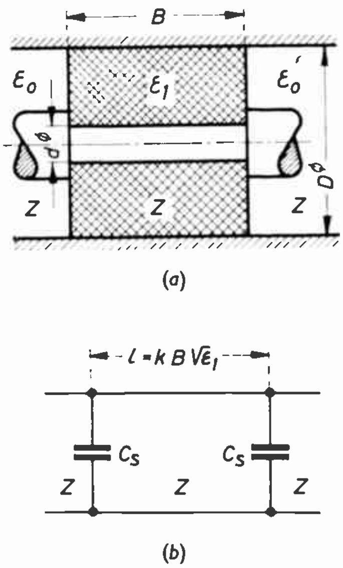


Fig. 2. (*a*) Broad support, without compensation, of diameter ratio

$$\frac{D}{d} = \exp\left(\sqrt{\varepsilon}\frac{Z}{60}\right)$$

(*b*) Its equivalent circuit.

The measured reflection coefficient of a test item with $Z = 60\,\Omega$, $B = D = 58$ mm, and $\varepsilon = 2{\cdot}55$ is shown as a function of frequency in Fig. 3. The abscissa is the product of frequency and outer-conductor diameter—$f \times D$ in Gc/s and cm. As described in the first part of this paper,[1] high measurement accuracy has been obtained by the node-shift method. The same measuring equipment has been used. A dotted curve is given in Fig. 3 for comparison. It refers to the equivalent circuit of Fig. 2 (*b*) and has been determined by calculation, the spurious capacitance C_s being assumed as 0·18 pF and the length l a value equivalent to the test item. The electrical length $l_e = B \times \sqrt{\varepsilon_1}k$ is longer than the geometrical length l by the factor k and depends on the dielectric constant ε, because a spurious electric field is present at the ends of the support. In the arrangement used, k is approximately 1·02.

A comparison between the calculated and measured curve shows good agreement, in particular at low frequencies. Only where the electrical length is greater than $\lambda/8$ do the curves appreciably differ from each other. The greatest difference occurs at the point $\lambda/4$, where the reflection curve of the equivalent circuit naturally goes to 0, while the measured curve exhibits a measurable residual reflection ρ_r. Zero reflection of the equivalent circuit at $\lambda/4$ results from the phase opposition of the spurious amplitudes which are to be added, under the condition that the equivalent spurious capacitances C_s are lumped. The residual reflection shown by the measured curve is caused by the fact that the spurious fields at the edges of the support have a finite extension and the lines of force are bent symmetrically to each other. They are passed through by the transmitted and the reflected wave in opposite directions. Because of this asymmetry the reflection is not fully compensated and a small residual reflection ρ_r remains. This disturbing effect is not great. For example, in the measured curve of Fig. 3 ρ_r is 0·3%. The equivalent circuit of Fig. 2 (*b*) is therefore a good approximation for the support, or, in other words, the discontinuity is caused by two equal capacitances located at the edges of the support.

3. Determination of the Dielectric Constant by Plotting the Reflection-Coefficient Curve

A measurement of the reflection coefficient of broad supports (Fig. 3) as a function of frequency for different dielectrics, e.g. $\varepsilon = 2{\cdot}05$, $2{\cdot}55$, $3{\cdot}55$, shows that the reflection at the $\lambda/4$ point is a minimum when the diameter d in the support exactly complies with relation (2). It is typical that small departures from the optimum diameter d cause appreciable changes of the reflection coefficient. Thus it is possible, by varying the diameter d until the residual reflection ρ_r is a minimum, to calculate, from the diameter d_z

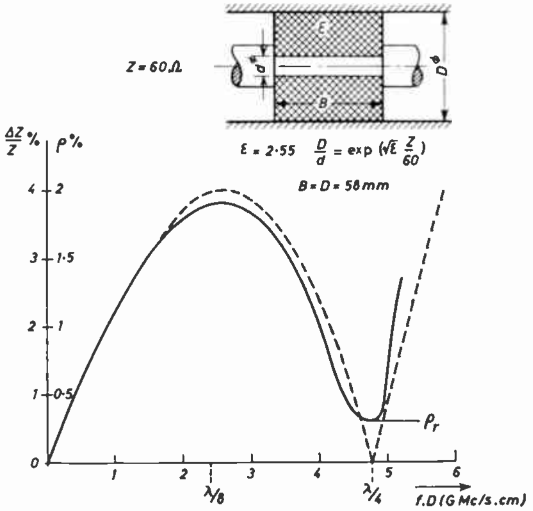


Fig. 3. Measured reflection coefficient of the broad support of Fig. 2 (*a*) as a function of frequency (solid curve); calculated reflection coefficient of the equivalent circuit of Fig. 2 (*b*) (dotted curve).

thus obtained, the dielectric constant ε. Thus

$$\varepsilon = \left(\frac{60}{Z}\right)^2 \left(\log_e \frac{D}{d_z}\right)^2 \qquad \ldots\ldots(3)$$

$$\rho_r = \text{optimum}$$

This method permits very accurate determination of the dielectric constant ε. One obtains, for example, for teflon 2·05, for trolitul 2·55 or 2·53. These values are in good agreement with the results obtained by other authors.[4, 5]

In this connection it may be added that the low-reflection supports described in the following can be designed only if the ε value is accurately known. Since according to past experience the measurement of the dielectric constant of a plate capacitor gives too low a value at low frequencies, such a measurement would be too inaccurate for the present purpose. Values of adequate accuracy at u.h.f. and v.h.f. are obtained by measurements using a coaxial line[5] or a cavity resonator.[3, 4] Also suitable is the method described above, determining the diameter ratio D/d where the residual reflection is a minimum.

Usually trolitul is used for supports; its dielectric constant depends on the type. For trolitul VI $\varepsilon = 2{\cdot}53$[3], for trolitul EF $\varepsilon = 2{\cdot}55$. Variations of dielectric constant for the same type have not been found. The dielectric constant of teflon varies between 1·98 and 2·05, depending on the manufacturer.

4. Compensation of the Spurious Capacitance at the Point where it Originates

As explained above, the spurious capacitances can be located at the edges of the support. To keep reflection low, they should be compensated immediately at the point where they originate, i.e. at the edges of the support, rather than by an inductive line section in the centre of the support. A few simple technical solutions are shown in Fig. 4. In the example (i), the inner conductor of diameter d projects by a small length a from the support. The absence of the dielectric here creates a zone of increased inductance, which compensates for the spurious capacitance originating in close proximity. Most of the inhomogeneous lines of force originate from the vertical face of the enlarged inner-conductor cross section and the compensating inductance is also located there. Moreover, the air dielectric reduces the spurious capacitance C_s, because most of the inhomogeneous lines of force now run through air. Even if the length a_1 (Fig. 4 (ii)) is filled with dielectric, the inductance is sufficient for compensation and the spurious capacitance C_s is reduced. This compensation may therefore be very effective.

For many practical purposes the support must be protected against sideways displacement. A suitable type of support is shown in Fig. 4 (iii). The dielectric penetrates into the outer conductor so far as is necessary for mechanical stability; the ratio $D'/D = 1{\cdot}05$. The characteristic impedance in the support depends on the large diameter D'. Therefore, the condition is $D'/d = \sqrt{\varepsilon}\,\frac{Z}{60}$. The compensation length a_1 of the supports of Figs. 4 (ii) and (iii) must theoretically be greater than a, because the air space is here filled with dielectric. Its practical influence is small because the dielectric goes over into air at the diameter of the inner conductor and the dielectric constant of the series arrangement is not appreciably greater than 1.

5. Optimum Dimensions Determined by Measurement

Measurements were made with the three most important characteristic impedances, $Z = 50$, 60, 75 Ω, and the dielectric constant $\varepsilon = 2{\cdot}05$, 2·55, 3·55. The compensating length a at the edges of the support (Fig. 4 (i)) was varied until the optimum reflection coefficient was reached as a function of frequency. A few typical examples of the great number of measured curves are shown here (Figs. 5, 6 and 7). The reflection-coefficient curves are similar for all supports of this kind. The abscissa is again $f \times D$ (Gc/s × cm); the figures refer to Gc/s for a line having an outer conductor of 1 cm diameter. For the practical support, the frequency is obtained by dividing by the outer-conductor diameter D.

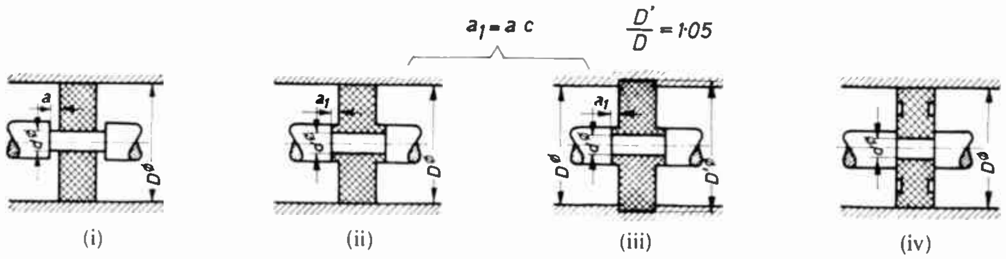


Fig. 4. Four supports of diameter ratio $\frac{D}{d} = \exp\left(\sqrt{\varepsilon}\,\frac{Z}{60}\right)$ with compensating lengths a, a_1 at the borders.

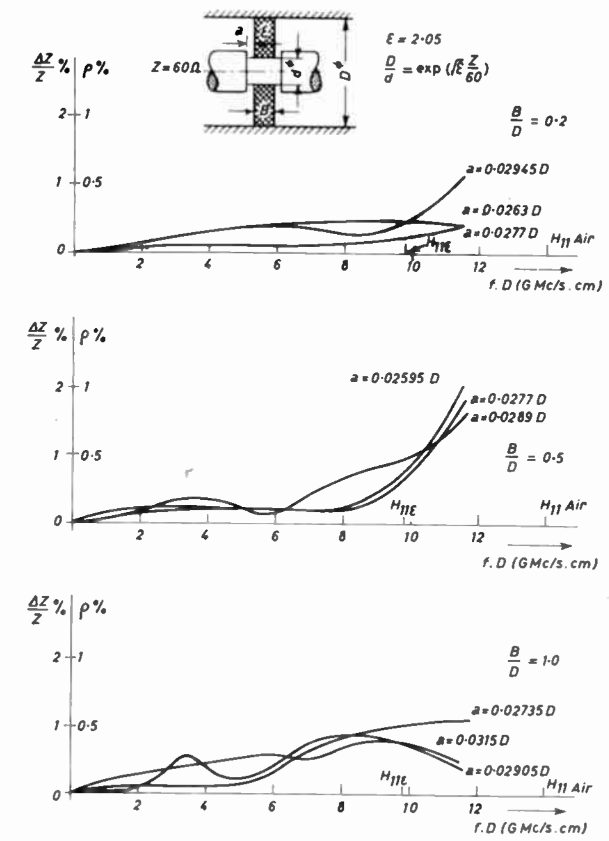


Fig. 5. Reflection-coefficient curves of supports with compensation (a) at the borders for various widths, with $\varepsilon = 2{\cdot}05$.

Comparison of the reflection coefficients of the new support and of the earlier one[1] shows that the new one is much better. For supports of medium breadth the factor of improvement is about 3. This shows that it is important to compensate for the spurious capacitance directly where it arises. For teflon and trolitul, the reflection coefficient is so small that operation at a frequency higher than the limit of the H_{11} wave (Figs. 5, 6) in the support is possible. The H_{11} value marked on the abscissa is true for the relation

$$\lambda_k \simeq \frac{\pi}{2}(D+d)\sqrt{\varepsilon} \qquad \ldots\ldots(4)$$

Excitation of the H_{11} wave (Fig. 8) is not likely because of the symmetrical configuration; but it should be borne in mind that the reflection coefficients here involved are extremely low and may be caused by a H_{11} wave of very small amplitude. It is found, however, that the operating frequency may even be approached to the limit of the H_{11} wave in the air line, which is higher by $\sqrt{\varepsilon}$ than in the support, without causing too large reflection. Surprisingly poor values are obtained for the support with the high dielectric constant $\varepsilon = 3{\cdot}55$.

It is typical that the supports (Fig. 4) have the same characteristic impedance as the adjoining line. Since the inhomogeneous fields at the edges of the support hardly penetrate into the support, the middle section is homogeneous. It may be prolonged without reaction on the inhomogeneous fields. The supports compensated at the edges may have any desired breadth without influencing the reflection at the edges. The resulting reflection is the vector sum of the two edge reflections; for very broad supports, it may describe an undulating curve, but this case seldom occurs. Usually, as described earlier,[1] one obtains a simple low-pass filter curve. The compensating lengths a and a_1, the limit values for a very broad support, hold also for a single transition from the dielectric to air (Fig. 9), which takes place from a cable with solid insulation to one with air insulation.

It was found by numerous measurements that the compensating length a is virtually independent of

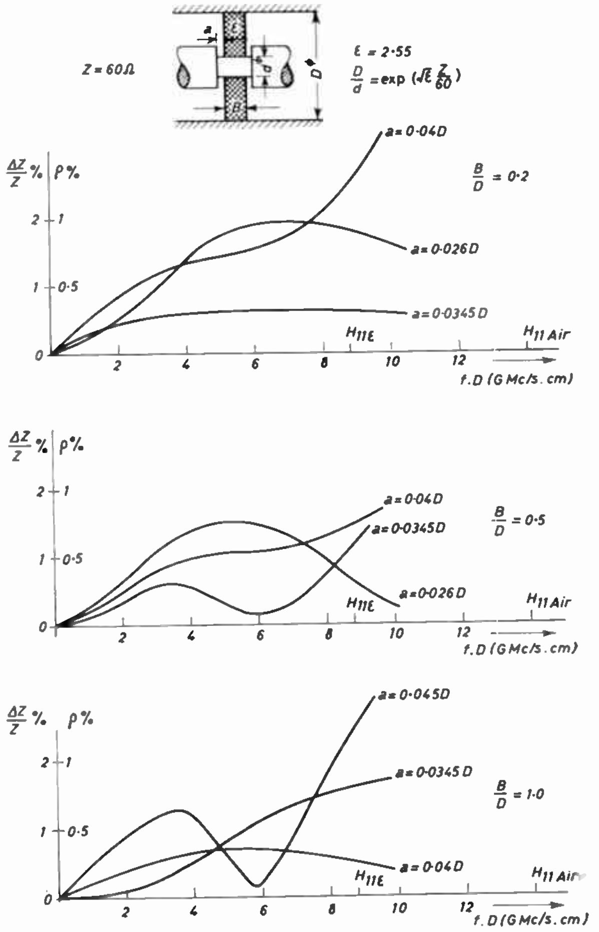


Fig. 6. Reflection-coefficient curves of supports with compensation (a) at the borders for various widths, with $\varepsilon = 2{\cdot}55$.

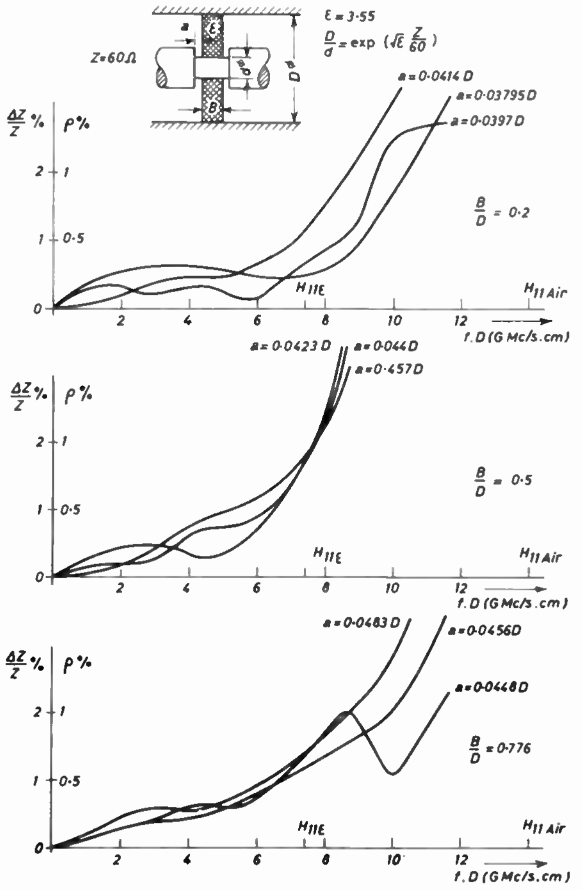


Fig. 7. Reflection-coefficient curves of supports with compensation (*a*) at the borders for various widths, with $\varepsilon = 3{\cdot}55$.

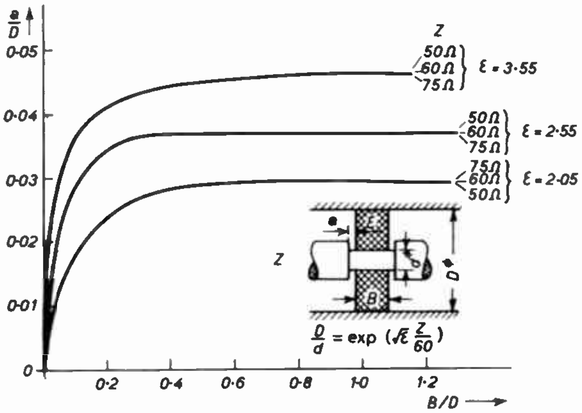


Fig. 10. Optimum dimensions of supports compensated at the borders, as a function of width *B*, ε being parameter.

the characteristic impedance. The curves of Fig. 10 give optimum dimensions for all characteristic impedances between 50 and 75 Ω.

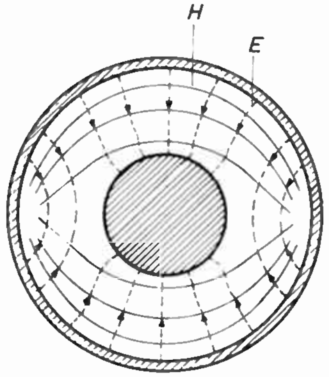


Fig. 8. Field configuration of the H_{11} wave in a coaxial line.

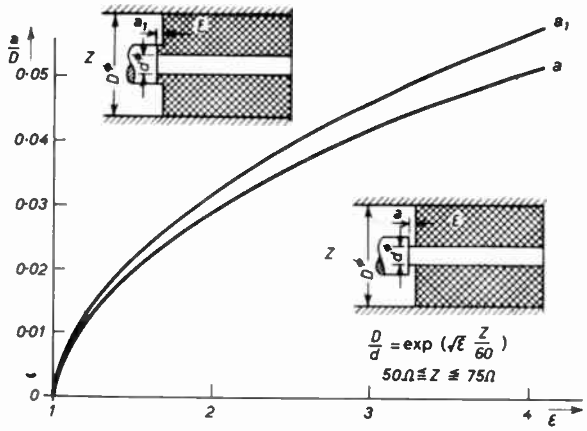


Fig. 9. Optimum dimensions of the dielectric-to-air transition as a function of the dielectric constant.

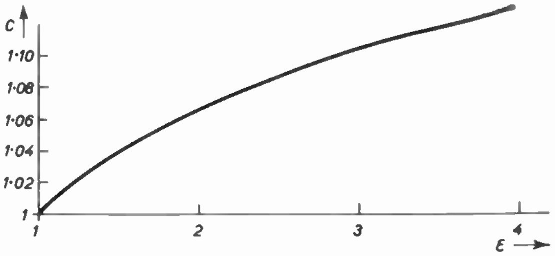


Fig. 11. Correction factor *c* for supports compensated at the borders according to Figs. 4 (ii) and (iii), as a function of the dielectric constant.

The compensating length a_1 of the supports of Figs. 4 (ii) and (iii) is somewhat greater than the first compensating length *a*. The relation between a_1 and *a* is given by a factor *c*, depending on ε:

$$a_1 = a \times c \qquad \ldots\ldots(5)$$

The relation between *c* and ε is determined experimentally and plotted in Fig. 11.

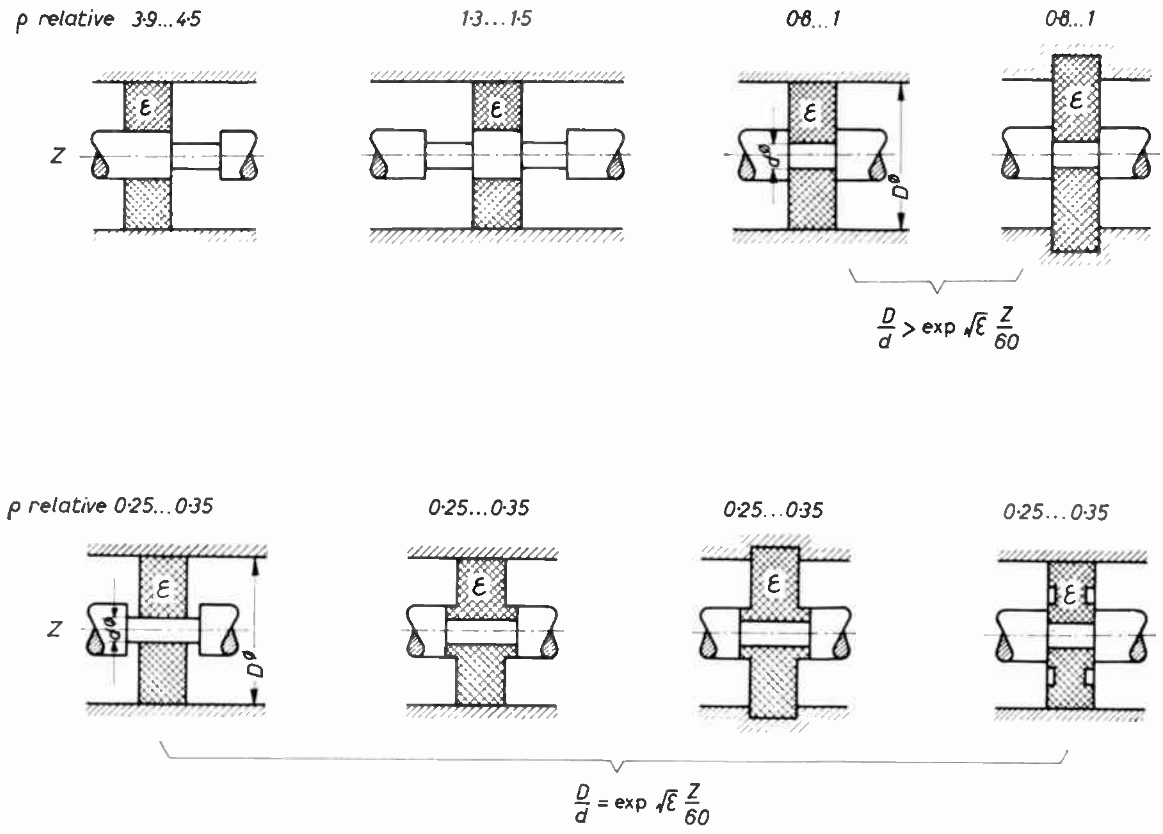

p relative 3·9...4·5
1·3...1·5
0·8...1
0·8...1
ε
Z
D
d
$\frac{D}{d} > \exp\sqrt{\varepsilon}\,\frac{Z}{60}$
p relative 0·25...0·35
0·25...0·35
0·25...0·35
0·25...0·35
$\frac{D}{d} = \exp\sqrt{\varepsilon}\,\frac{Z}{60}$

Fig. 12. Survey of supports of different shapes investigated to date.

Figure 12 summarizes the relative reflection for all supports investigated to date. It is evident that the supports compensated at the edges present substantial advantages over the others.

6. Axial Dimension of the Inhomogeneous Field

If several coaxial elements are arranged in close proximity, the resulting reflection of the complete arrangement is the vector sum of the individual reflections unless the individual inhomogeneous fields interfere with each other. According to past experience, it is preferable to make optimum compensation for every individual element if a number of discontinuities are to be combined. For the complete arrangement, the reflection may, in the most adverse case, be the arithmetical sum of the individual reflections. However, when using this method, care must be taken to ensure that the individual discontinuities do not influence each other when assembled, i.e. that they are spaced apart far enough in the axial direction. On the other hand, the individual elements must be disposed as close to each other as possible, if compactness is essential. For this reason, the limits of the inhomogeneous fields in the axial direction are of interest.

The curves shown in Fig. 10 become horizontal at a certain breadth B. The compensating length a becomes independent of B. This shows that the inhomogeneous fields, which are compensated by a, no longer influence each other. Since the inside of the support has a homogeneous field and the correct characteristic impedance, this section may be extended arbitrarily without affecting the boundary fields. The limit where the inhomogeneous fields end is indicated by the point where the curve becomes horizontal. This length is here called l_g. Since the inhomogeneous fields act on both sides of the support, l_g is half the amount of the critical breadth where the curve turns to the horizontal. With $\varepsilon = 2{\cdot}05$ and $\varepsilon = 2{\cdot}55$, the critical length $l_g \simeq 0{\cdot}20\ D$, with $\varepsilon = 3{\cdot}55$ we find $l_g \simeq 0{\cdot}4\ D$. If this discontinuity is followed by a second one, the critical spacing l_{g1} must be observed, while for the second discontinuity, of course, the corresponding spacing l_{g2} must also be kept. The distance between two discontinuities is therefore the sum of the two critical lengths $l_{g1}+l_{g2}$. If sharp discontinuities, e.g. large changes in cross-section, are brought together, the distance to be observed between them according to past experience should at least be $0{\cdot}8\ D$ to ensure that each one is

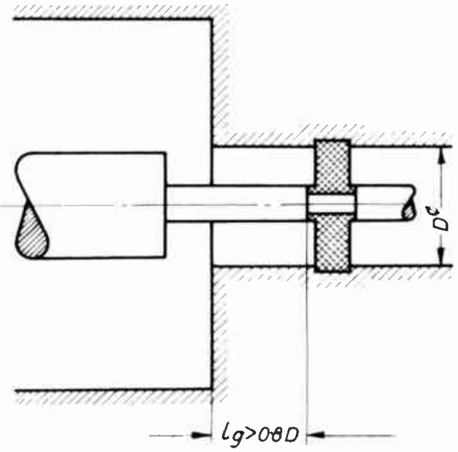


Fig. 13. Compensated cross-section transition with support axially displaced by $l_g > 0{\cdot}8\ D$ to avoid mutual disturbance.

effective individually (Fig. 13). Since the spurious fields decrease exponentially, an error arises if the distance is slightly shorter than the critical value. On the other hand, the mutual influence is sure to be negligible if the distance is made 1·0 *D*.

7. References

1. A. Kraus, "Reflection coefficient curves of compensated discontinuities on coaxial lines and the determination of the optimum dimensions [Part 1]", *J.Brit.I.R.E.*, **20**, pp. 137–52, February 1960.
2. J. R. Whinnery, H. W. Jamieson, and T. E. Robbins, "Coaxial-line discontinuities", *Proc. Inst. Radio Engrs*, **32**, pp. 695–709, November 1944.
3. M. Gevers, "Measuring the dielectric constant and the loss angle of solids at 3000 Mc/s", *Philips Tech. Rev.*, **13**, No. 3, pp. 61–70, September 1951.
4. F. Gross, "Temperature dependance of loss angle and dielectric constant of solid insulating materials in the range about 4000 Mc/s", *NachrTech. Z.*, **9**, pp. 124–7, March 1956.
5. R. Eichacker, "A material-characteristics test assembly for determining the electromagnetic material constants of solid and liquid media at frequencies between 30 and 7000 Mc/s and temperatures between − 60 and + 240°C", *R & S-Mitteilungen*, No. 11, pp. 185–205, 1958. (Original in German—English reprint available.)

Manuscript received by the Institution on 29th July 1961. (Paper No. 726.)

News from the Sections . . .

Southern Section

At very short notice, because of the illness of the advertised speaker, Mr. A. E. Crawford (*Member*) gave an informal paper entitled "Force-Electric Generators" on 10th January. Mr. Crawford began by pointing out that the requirement for sources of high voltage or high current in the form of pulses has increased recently, due to the demands in nuclear research, radar equipment and high power flash tubes. In the past, the problem has mainly been approached using large banks of non-inductive capacitors. Developments in the field of ferro-electric materials have shown that the inherent energy contained in a polarized ferro-electric can be released by a mechanical shock wave of an intensity sufficient to cause a return to the random domain orientation state. The direct piezoelectric effect in these materials is also capable of producing high voltages without destruction of polarization. Mr. Crawford outlined the principle of operation of both methods of generation and illustrated his talk with the aid of a number of slides.

J. M. P.

North Western Section

On the 1st March at the College of Science and Technology, Manchester, Mr. P. Lowry (Associate Member) read a paper entitled "A Colour Television Projector for Medium Screen Applications".

Mr. Lowry, in his introduction, stated that since the very early days of television there has been a desire to present pictures in colour but, as in photography, economic and engineering limitations have imposed formidable restrictions on the rate at which colour reproduction has developed. In the home, the demand is for a compact, inexpensive, easily mass-produced display giving a sufficiently bright picture with normal room lighting and a picture of adequate size. Similar requirements exist for closed circuit installations but for large audience viewing for entertainment or instruction then a large area is preferable. For monochrome television a range of monitors exists which satisfy group requirements and, for bigger audiences, medium and large screen projection displays are available.

After describing the basic requirements for the simultaneous transmission of colour by television whereby the pick-up device scans the picture and produces the three signals representing the red, blue and green components the lecturer explained how projection systems could be arranged and optical systems positioned. Detailed requirements were then discussed, including choice of cathode ray tube, optical systems, correction for the various forms of distortion and the scanning and video drive arrangements. Methods of operating were mentioned and the mechanical construction described. Finally, Mr. Lowry gave a summary of performance and application.

F. J. G. P.

Scottish Section

The recent steady increase in attendance was well maintained when 40 members and guests were present at Edinburgh on 7th February and 66 at Glasgow on 8th February. The speaker at both meetings was Mr. A. E. Hilling, B.Sc.(Eng.), who read a paper on "Transistors in Communication Transmitters and Receivers".

After briefly outlining the development of the transistor up to the alloy diffusion and mesa types, he described what could now be achieved with them at h.f. and v.h.f. New thinking was necessary, he said, in developing circuitry to deal with the special problems encountered, particularly when frequencies up to 800 Mc/s were considered. A major difficulty was the dependence of transistor parameters on temperature, and he outlined methods by which this could be reduced.

In proposing a vote of thanks to the speaker, Mr. E. G. W. Miller said he felt acutely aware of the fact that we had now entered a new era in which it was no longer strange for an electronic engineer to claim that he knew more about transistors than thermionic valves.

W. R. E.

South Western Section

On the 14th March, the South Western Section held a joint meeting with the British Computer Society at the Bristol College of Science and Technology to hear a paper by Dr. A. D. Booth (*Member*) entitled "Recent Applications of and Refinements in Electronic Digital Computers". The South Western Section Chairman, Mr. H. H. Harper, welcomed the members of the Society and introduced the lecturer.

Dr. Booth opened with a warning of the dangers of the uncritical acceptance of preconceived ideas, and then proceeded to demolish a number of common misconceptions regarding Boolean algebra, time-sharing, time-of-carry, and storage devices. Some recent applications were then discussed, including auto-coding systems and machine translation.

The discussion which followed was opened by Mr. J. M. Hahn for the British Computer Society and continued briskly, covering every aspect of Dr. Booth's paper. The large attendance at this meeting bore witness to the authority of the lecturer and the interest which his subject had aroused among members of the British Computer Society as well as those of the Institution.

G. F. N. K.

U.D.C. 534.23: 620.179

The Physical Factors Affecting the Reliability of Ultrasonic Non-Destructive Testing

A Review of Current Research

By

L. KAY, B.Sc.,†

E. WHIPP, B.Sc.,†

AND

M. J. BISHOP, B.Sc.†

Presented at the West Midland Section's Symposium on "New Electronic Techniques in Non-Destructive Testing" in Wolverhampton on 6th December 1961.

Summary: The formation of a beam and the deflection of this relative to a boundary is being investigated to determine the reliability with which one can specify the radiation of stress waves in a solid. Ultimately it is hoped to produce an electronically scanned beam which will be free from transverse waves. Little information is available regarding the propagation of stress waves in metals and much of the literature gives conflicting results. Before the frequency of operation of an ultrasonic testing device can be decided, the metallurgical characteristics of the material should be related to the scattering and absorption which may be expected over a wide frequency range. The means for doing this are not available and an attempt is being made to provide them in a suitable form. Signal processing seems to play a very small part in ultrasonic testing. Nothing appears to have been attempted to increase the probability of detection of a defect when viewed against a background of spurious returns. The problem is made difficult when compared with radar or sonar since the medium is stationary, but means are available theoretically whereby the background due to a large number of small signal returns can be made to change relative to an echo from a larger discontinuity in a stationary medium. This is being investigated experimentally.

1. Introduction

Since high-frequency stress waves were first used for detecting defects in metals, considerable progress has been made in the technology of ultrasonic non-destructive testing, and commercial equipments are so versatile that they can be applied to many of the day-to-day problems of the metallurgical industries as well as routine testing programmes. Even so, ultrasonic testing of materials is still not utilized on a large scale in many industrial organizations because of the unreliability of some of the results. It has been shown by Claydon,[1] for example, that the size of a defect may be very different from that indicated by ultrasonic tests. This is partly due to the various shapes encountered in defects, and is partly due to the non-uniform stress field in the material, as shown by Christie.[2] It is in fact generally acknowledged that ultrasonic non-destructive testing is more of an art than a science depending very largely upon the experience and skill of the operator of the equipment.

A critical study of the basic principles of ultrasonic non-destructive testing reveals there are more fundamental uncertainties in the method than would normally be acceptable in most other forms of industrial measurements. Nevertheless, no other method can stand comparison. Many papers on the subject liken the system to that of sonar; they do not mention, however, that sonar is far from being an infallible means of detecting underwater objects—albeit the best. All echo-location methods suffer from a degree of uncertainty because of the inhomogeneity and absorption characteristics of the medium in which the energy is propagated, the complexity of the shape of the objects to be detected and examined, and the coarse resolution in both range and direction resulting from the relatively long wave-length employed—as compared with that of light. Metals are among the worst of mediums for propagating sound waves, and many of the new materials on which testing is of an ever-increasing importance show a further deterioration from the ideal ultrasonic characteristics. How then can we hope even to keep pace with metallurgical development, much less advance ahead of it?

Research in this field of non-destructive testing has been on a very limited scale and quite uncoordinated; and much of the knowledge has been gained from studies of the ultrasonic delay line as used in communications systems. The classical books on stress waves in solids make only a cursory mention of the

† Electrical Engineering Department, University of Birmingham.

practical applications to flaw detection. The reasons are all too obvious.

The purpose of this paper is to discuss some of the outstanding difficulties and explain how these are being investigated.

2. Problems Outstanding in Ultrasonic Non-Destructive Testing

Almost all the problems in non-destructive testing can be grouped under one heading "uncertainty", but it is the uncertainty of so many factors that make this particular problem both formidable and interesting. There is the uncertainty of detecting a defect, the uncertainty of its size, shape, and nature, and finally the uncertainty of its importance. Even this latter is so often arbitrarily stated, based on experience alone and often without reason.

These uncertainties would be reduced to insignificance if the operator of equipment were presented with the equivalent of an optical picture, but to do this we would have to use frequencies greater than 100 Mc/s so that the wavelength is very small compared with the irregularities of the object. Such frequencies appear to be impractical since energy at 10 Mc/s and above will not propagate through a reasonable thickness of many metals. When the question is asked, "What are the factors affecting the propagation of stress waves in solids composed of anisotropic crystals?" we find in fact that no satisfactory answer is available—the information on the subject of attenuation for example is conflicting—and a new approach seems to be required.

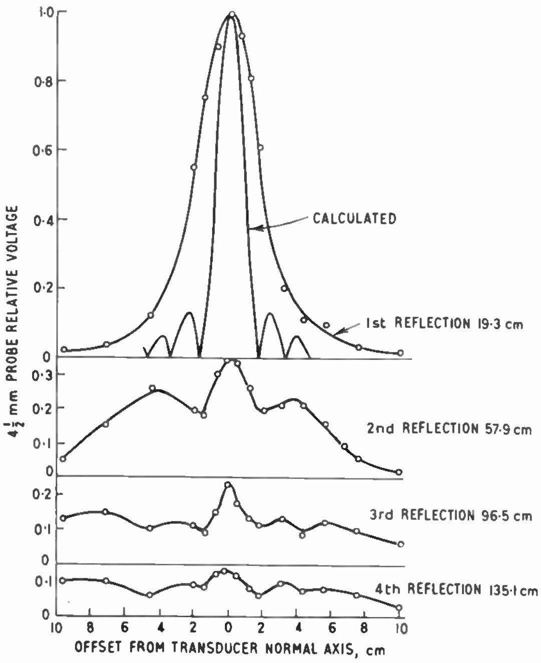


Fig. 1. Direct field pattern and reflection field patterns for 1·75 Mc/s. 1 in. diameter plain disc measured in a block of aluminium 19·3 cm thick.

The formation of a beam, whether it be diverging or focused, in a metal is understood even less. Christie[2] made the first constructive approach and showed that in a liquid, which will support only longitudinal waves, the ultrasonic field produced by a simple disc transducer was much more complex than was at first thought to be the case. If we are to analyse the signal received by an ultrasonic transducer it is necessary that we must know the nature of the field producing the signal. This can only be done by first knowing the ultrasonic field of the source of energy. The literature shows there is considerable scope for both experimental and analytical study.

Finally, the certainty with which a decision can be made about any one signal among a background of similar signals can only be improved by some form of signal processing. By this is meant the extraction of the maximum amount of information contained in the signal from the transducer. Usually some form of correlation technique is involved. Examples of this can be found in the radar field, such as frequency-modulation systems including "chirp" radar, multi-pulse systems, multiplicative and additive arrays, and integration methods. They all probably involve some form of correlation process and such processes are not in use in present-day ultrasonic instruments. It would seem reasonable that some improvement may be possible if they were applied to the problem.

Compared with sonar or radar, research along these lines has been almost non-existent and may pay dividends if exploited. This at least is the approach adopted at the University of Birmingham and the work there will now be discussed.

3. Ultrasonic Propagation in Solids

The intensity of stress waves in the far-field region of an ultrasonic transducer is assumed to follow the simple law of spherical spreading modified by an attenuating factor, namely

$$I_r = \frac{I_0 G_{(\theta)} \exp(-\alpha r)}{4\pi r^2} \qquad \ldots\ldots(1)$$

where $\frac{I_0}{4\pi}$ = intensity at unit radius if the source were a point radiating into 4π radians.

r = distance from the source.

$G_{(\theta)}$ = directivity function relating the intensity at an angle θ to the axis with that from an isotropic source.

α = attenuation coefficient which may be complex.

The minimum distance at which we can assume such an expression to be valid is often quoted as being the limit of the Fresnel region. In systems such as sonar this is usually acceptable, since most objects of interest are well into the Fraunhofer region. Those which are not, are at least comparable in size or greater than the transducer dimensions. Ultrasonic testing of metals, on the other hand, usually involves the Fresnel and Fraunhofer regions and the defects to be examined are generally smaller than the transducer dimensions; hence the analogy with sonar breaks down. Ray theory cannot in fact be applied satisfactorily and a detailed study of the wave equations is necessary to determine the ultrasonic field. Such a study usually demands certain simplifying assumptions and the alternative approach is an experimental investigation of the propagation characteristics, i.e. beam formation, the scatter of energy by the crystal structure and by inhomogeneities, and absorption under realistic conditions.

The ultrasonic field of a circular disc transducer has therefore been explored in both the "near" and relatively "far" fields using large diameter cylindrical blocks of aluminium. The block for examining the "far" field was 19·3 cm thick and 23 cm diameter. The transducer was 2·54 cm diameter and was driven at about 1·7 Mc/s producing a near field region of the order of 4·5 cm. A probe of 4·5 mm diameter was used initially for measuring the field intensity on the opposite face of the block to that on which the transmitting transducer was placed. As the probe was traversed across the flat face, the voltage generated in the probe followed the curves of Fig. 1. The measurements were obtained using the first pulse to arrive at the probe, then subsequent pulses due to multiple reflections at the flat parallel faces.

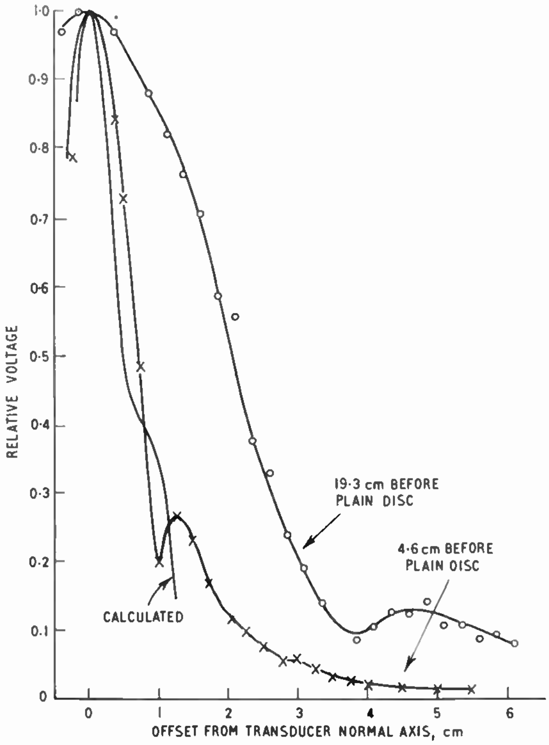


Fig. 2. Transducer field patterns measured in a block of aluminium 4·6 cm thick compared with a calculated pattern and that measured in a block 19·3 cm thick.

Comparing these curves with the theoretical curve for a circular disc transducer obtained from the relation

$$\frac{P_\theta}{P_0} = \frac{2J_1\left(\frac{\pi d}{\lambda}\sin\theta\right)}{\frac{\pi d}{\lambda}\sin\theta} \qquad \ldots\ldots(2)$$

where

P_0 = pressure amplitude on the axis,
P_θ = pressure amplitude at an angle θ to the axis,
d = diameter of the disc,
J_1 = Bessel function of the 1st order,

we see that there is disagreement. The main lobe is wider than that calculated, and there are no points of zero amplitude. This latter feature is partly accounted for by the size of the probe. After reflection, the peak of the main lobe is very much reduced relative to its side lobes, and this is largely due to the energy extracted by the transmitting transducer each time reflection takes place. A smaller probe of diameter 1·2 mm was used to obtain the experimental curves shown in Fig. 2, where a minimum was observed. Equation (2) predicts zero amplitude but is applicable only to the true "far" field of a transmitter. However, at a distance of 19·3 cm, a minimum amplitude smaller than that obtained was expected. The measurements were made with short pulses instead of with continuous wave transmissions. Transient effects in the build-up of the ultrasonic field and in the transducers could be seen and measurements were taken only when c.w. conditions had been established within the pulse. Effects due to the finite size of the probe at the position of the minimum were considered insignificant since the phase error across the probe diameter was less than 3 deg.

Non-piston-like vibration of the transmitting source could account for these unpredicted features. The amplitude of vibration across the face of the transducer has been found by previous work[3] to vary considerably depending upon the ratio of diameter to thickness. Modes of vibration in a disc, other than a simple thickness mode, are simultaneously excited resulting in non-piston-like vibration in the thickness mode. At the edge of the transducer the amplitude of vibration of the metal face cannot immediately become zero but must taper off, thereby effectively increasing the diameter of the transmitting source. The increased diameter would be expected to produce a narrower beam, but it is suggested that because of

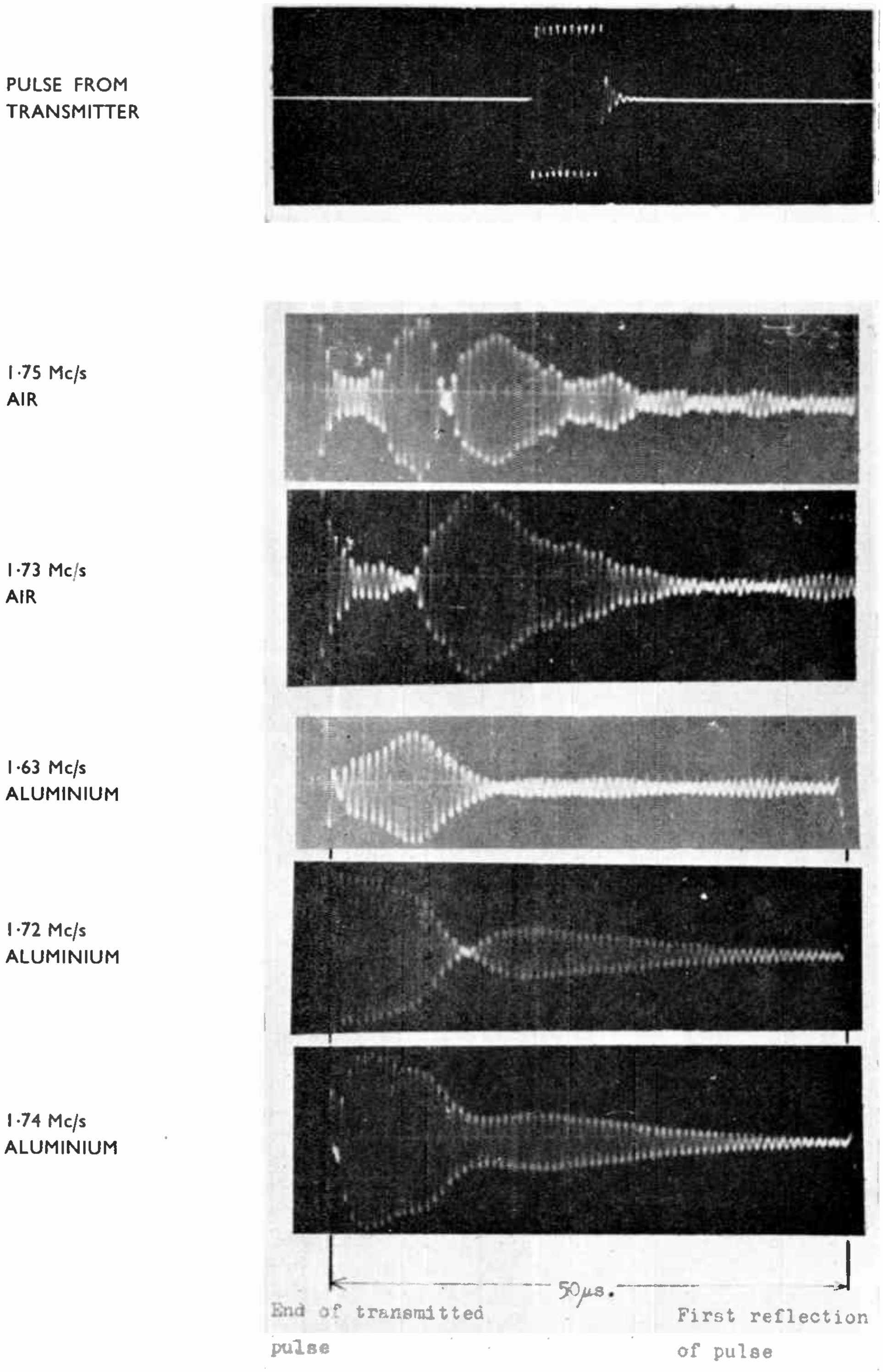


Fig. 3. Oscillographs of transducer voltage when used as a receiver following the transmission of a pulse.

the non-uniform amplitude of vibration of the source across its face, the beam width is increased and the side lobes affected. It is quite evident that before scattering and attenuation measurements can be made, more must be known about the influence of these effects on the formation of the beam.

Similar measurements on a 2·3 cm diameter cylinder of aluminium 4·6 cm thick gave the result also shown in Fig. 2 where it will be seen that the field intensity was again significantly wider than that calculated for that region of the "near field".

When the transmitting transducer was also used as the receiver, the signal received immediately following the transmission of a pulse was examined carefully to see if there was evidence of backscatter. The voltage waveform revealed the transducer to be vibrating in more than the simple longitudinal mode. Conversion from one mode to another was taking place with the result that a large amplitude signal was present for a considerable time after the cessation of the transmission driving pulse. Examples of this are shown in Fig. 3. Slight variation of the driving frequency caused large changes in the envelope of the signal. Before back-scatter can be measured with a common transmit-receive transducer this form of vibration must be stopped.

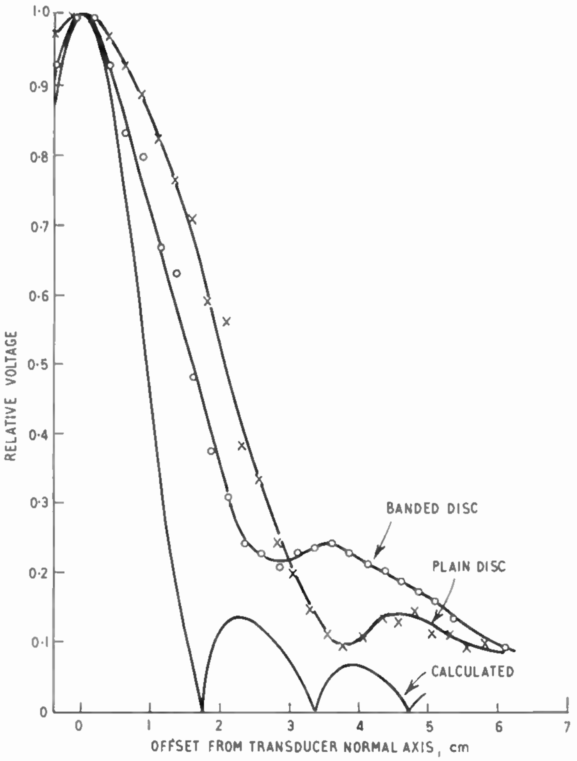


Fig. 4. Trɘnsducer field patterns for a plain circular disc and a disc having concentric grooves.

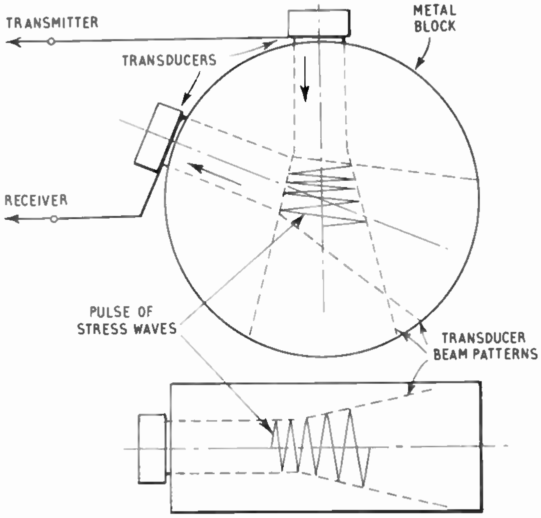


Fig. 5. Arrangement for angle-scatter measurements.

If various modes of vibration are taking place in the transducer,[3] the phase and amplitude of the thickness mode at points along a radius of its face will not be uniform. To reduce the coupling between radial modes of vibration and the thickness mode, a barium titanate disc was grooved with concentric rings. This was a first stage in making a transducer consisting of elements mechanically separated. Each element would be a point source of radiation, so that for equal electrical excitation of all the elements the transducer would tend towards a source with uniform amplitude and phase of vibration at all points. The measured beam pattern for the grooved disc is shown in Fig. 4 where the beam is narrower than that of a plain disc of barium titanate.

A new approach to the measurement of scattered signals is being developed using a cylindrical block having a transducer fitted to the circumference of the block. Direct transmission will be measured opposite the transmitting transducer, and angle scatter at points off the axis will be determined as shown in Fig. 5. Scattered signals have been observed by these means but before the results can be analysed the ultrasonic field of the transducers must be understood in more detail than at present.

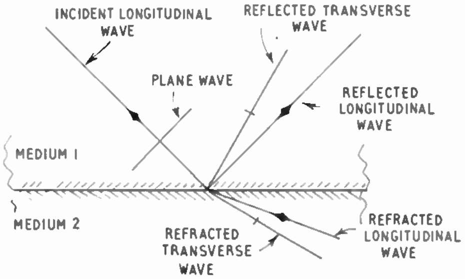

INCIDENT LONGITUDINAL WAVE
REFLECTED TRANSVERSE WAVE
PLANE WAVE
REFLECTED LONGITUDINAL WAVE
MEDIUM 1
MEDIUM 2
REFRACTED TRANSVERSE WAVE
REFRACTED LONGITUDINAL WAVE

Fig. 6. Reflection and refraction of stress waves in solids.

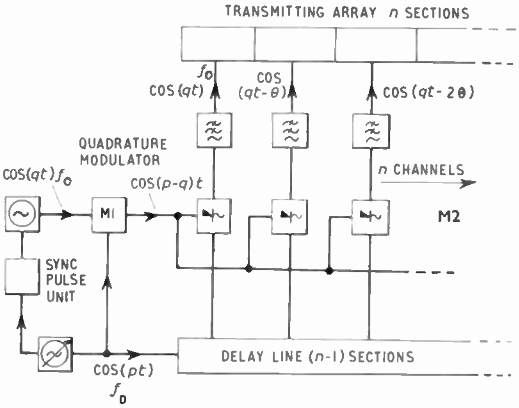


Fig. 7. Beam deflection system.

4. Electronically Scanned Beams in Metals

Several devices for causing an ultrasonic beam to be deflected from the normal to the surface of a metal have been proposed and some are in current use, but generally speaking they suffer from a lack of detailed knowledge about the formation of a deflected beam in the metal. A familiar diagram shown in Fig. 6 illustrates the reflection and refraction of stress waves at a boundary and this has been used as a means for postulating the direction of propagation of the stress waves generated by the devices. As an alternative to this method, the near field of a multi-element transducer is being examined in water to determine the limitations which must be put on the use of such a transducer, and the effect of applying a phase taper to deflect the beam from the normal. The equipment has now been completed and the system to be used is shown in Fig. 7.

An electromagnetic delay-line is used to feed the sections of the transducer via a frequency changer in each channel. The purpose of the delay line is to introduce a phase difference of equal amount between each section, which can be varied from $\theta = +\pi$ to $\theta = -\pi$ radians. This is achieved by applying a frequency which can be varied from $f_D = f_1$ to f_2 corresponding to a delay per section of $-\pi$ to $+\pi$ radians respectively. The frequency to be transmitted must be constant, and this is arranged by means of the modulators M1 and M2. The lower side-band of the single side-band modulator M2 has a frequency $f_D - f_0$ which on being modulated with f_D from the delay line, gives $2f_D - f_0$ and f_0 etc. All frequencies except f_0 are filtered out before amplification and subsequent transmission.

Varying f_D thus causes a linear phase taper to be applied to the transducer array, since phase relations are maintained through the frequency changers. The effect is to deflect the direction of propagation as defined by the plane in which the particle velocity is in phase.

Referring to Fig. 8 it is seen that each element can be considered to have a near field and it is only when these completely overlap that the deflection of the beam starts to have some meaning. An undeflected beam will have regions of maximum and minimum particle velocity within the boundaries of the "beam" near the transducer face; these variations will also be present when a phase taper is applied. Using rectangular transducers will however reduce the variation as compared with circular transducers as shown by Freedman.[4]

An interesting feature arising from the use of a sectionalized array to generate stress waves in a solid is the effect this may have on the generation of transverse waves when a phase taper is applied to deflect the direction of propagation. The condition existing in the case of refraction at a boundary between two media is that the particle movement is at an angle to the boundary plane. A transverse wave is thereby generated, so that the stresses at the boundary will conform to those physically possible. An array of transducers vibrating longitudinally produces entirely different conditions and at present it is difficult to see how comparable transverse waves will be produced, but quite clearly the conditions are complex and are not readily amenable to mathematical analysis.

Bradfield[5] developed a system for deflecting a beam in discrete steps, but no results have been published.

5. Pulse and Frequency-Modulation Systems for Detecting Defects

The problem of detecting defects in metals is complex and one cannot hope to cover the many aspects met with in practice; only a general picture can be presented. For example, in fine grain metals, defects which have any metallurgical significance may

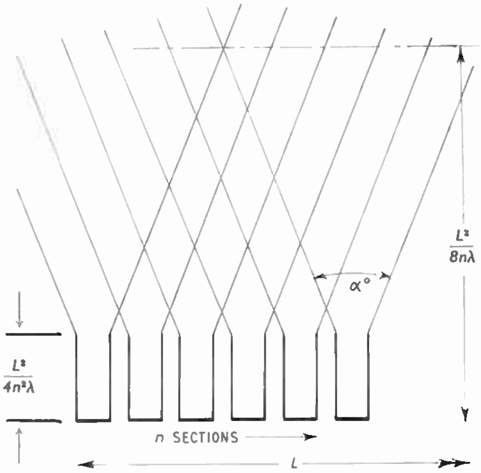


Fig. 8. Beam formation in near field of a sectionalized array.

be much smaller than in the coarse grain metals. A defect of any significance in a coarse grain metal would be easily detected if of the same size in a fine grain metal, and it is usually the defect which gives an echo of comparable size to that of the scatterer which causes difficulty. This is often stated in another way —the energy will not penetrate the medium due to scattering of ultrasonic waves at crystal boundaries and the depth at which a defect can be detected is very small.

There is insufficient evidence to show which of two factors predominate when a coarse grain material is used. The energy is either scattered so much that any resemblance to a beam is completely lost, or the sum of the scattered signals is greater than the signal from the defect. This latter case is the one now to be discussed.

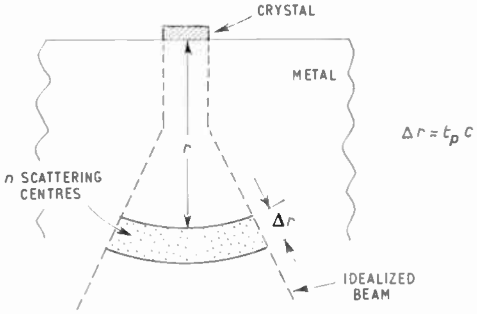


Fig. 9. Illustrating the increase in back-scatter with distance.

It can be assumed that the scattering centres in the metal are randomly distributed and are small in size. This applies equally well to coarse and fine grain metal since the crystal boundary is of random shape and has many scattering points. The signal received at the transducer at any one instant is the sum of all signals from an annulus in the metal at some distance as shown in Fig. 9. The intensity of the signal will decrease as $1/r^2$ because of the combined effect of spherical spreading (which we will assume for the present) and the increase in the number of scatterers as the distance increases. The signal from a defect on the other hand will decrease as $1/r^4$ so that the distance at which the scattered signal exceeds the defect signal will be much less for coarse grain metal than fine grain. We are however assuming the beam continues to be formed. To establish the validity of these assumptions, two experiments are envisaged:

(1) Effect of changing the frequency on the scattered signal and the defect echo signal.

(2) Effect of increasing the pulse duration of the scattered signal.

If, as we are assuming, there are a large number of scattering centres, a change in frequency will cause a change in the amplitude of the signal received from a particular annulus. It has been shown[6,7] that a change in frequency equal to (1/pulse-duration) causes the signal to be completely randomized. This means that the pattern of returns is no longer related to that obtained with the original frequency used. A defect echo on the other hand may not be affected to any great degree because it may be largely independent of frequency over the range of frequency change. Thus the background of signals due to the scatterers will vary, but not the defect echo. It must be remembered here that the echo, when observed, is the sum of the background signals and the defect echo, and as such the sum signal will also change.

We can illustrate this with probability distribution curves as shown in Fig. 10 which were obtained theoretically for noise and single tones.[8] The relation between band limited white noise and background signals from metals depends upon the number of scattering centres; if large, there will be a close statistical connection and the background signal will be unpredictable. Provided the bandwidth of the noise and the pulse producing the background signals is the same, their spectra will also be the same. Ideally the echo from a defect will be a replica of the transmission pulse, and if this is a short burst of tone, it can—for the purpose of statistical analysis—be considered as a tone signal whose amplitude is absolutely predictable. When the signal-to-noise ratio is large, as in the case for noise source N_1, the difference between the distributions is obvious and there is reasonable certainty of detecting the presence of the single tone. When the noise is large, as in the case of N_2, the chance of detecting the presence of a tone is very much less. These are the conditions pertaining in practice at present—one set of samples are given and a decision must be made from these. To improve

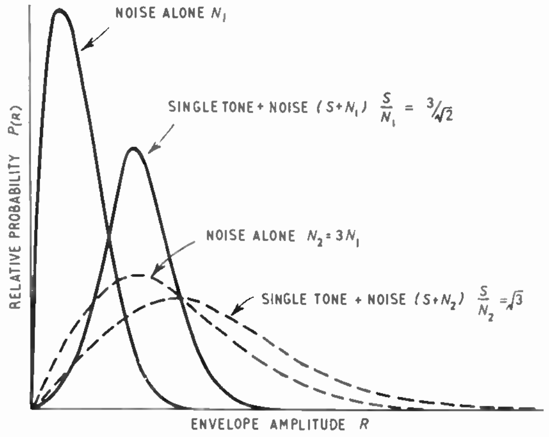


Fig. 10. Comparison between distributions for noise and noise + single tone.

the reliability of this decision a second set of samples is required, and this should be possible using a different frequency. An attempt to change the set of samples is made by operators of non-destructive testing equipment by moving the probe. This introduces other variables.

The improvement in reliability is related to the number of independent samples which are taken, and it is found that integration gives a theoretical improvement of 1·5 dB per doubling of the number of samples for low signal/noise ratios; subjective tests using either superposition of traces or side-by-side correlation gives an improvement of 2·4 dB.[9, 10] When the signal/noise ratio is greater than about 6 dB the improvement is 3 dB per doubling of the number of samples for both methods.

The problem of achieving these results in practice is being investigated and it appears the limit will be determined by the frequency band which can be achieved.

If now the pulse duration is halved, the volume of scatterers at any one point producing a signal will also be approximately halved. Since the scattered signals are assumed to add randomly, the sum signal will decrease by $1/\sqrt{2}$ times and since the echo signal amplitude will be unaffected an improvement in signal/noise ratio of 3 dB is possible. The combination of the results obtained by changing the frequency and the pulse duration will lead to a better understanding of the mechanism of the background returns against which a defect has to be detected.

6. Conclusions

We have shown that there remain many physical factors in ultrasonic non-destructive testing which affect the reliability of the results, and about which there is only very limited knowledge. The approach described in this paper should improve this situation, but by no means are all the problems being investigated and a more co-ordinated programme of research embracing the many research centres should be evolved if more rapid progress is to be made.

7. Acknowledgments

The authors wish to acknowledge the financial assistance given by the Department of Industrial and Scientific Research, and grateful thanks are due to Professor D. G. Tucker for the generous facilities which have been made available.

8. References

1. E. G. Stanford and J. H. Fearon, "Progress in Non-Destructive Testing", Vol. 2 (Heywood, London, 1960).
2. E. G. Stanford and J. H. Fearon, "Progress in Non-Destructive Testing", Vol. 1 (Heywood, London, 1958).
3. J. S. Arnold and J. G. Martiner, "Description of resonances of short solid barium titanate cylinders", *J. Acoust. Soc. Amer.*, **31**, p. 217, 1959.
4. A. Freedman, "Sound field of a rectangular piston", *J. Acoust. Soc. Amer.*, **32**, p. 197, 1960.
5. G. Bradfield, "Improvements in Ultrasonic Flaw Detection", *J. Brit.I.R.E.*, **14**, p. 303, 1954.
6. L. Kay, "A comparison between pulse and frequency-modulation echo-ranging systems", *J. Brit.I.R.E.*, **19**, p. 105, 1959.
7. L. Kay, "An experimental comparison between a pulse and a frequency-modulation echo-ranging system", *J. Brit. I.R.E.*, **20**, p. 785, 1960.
8. S. O. Rice, "Mathematical analysis of random noise", *Bell Syst. Tech. J.*, **24**, p. 46, 1945.
9. P. McGregor, "A note on trace-to-trace correlation in visual displays," *J. Brit.I.R.E.*, **15**, p. 329, 1955.
10. D. C. Cooper, "Integration and detection in pulsed radar and sonar systems," Ph.D. Thesis, University of Birmingham, September 1961.

Manuscript first received by the Institution on 13th November, 1961, and in revised form on 2nd April 1962. (Paper No. 727.)

U.D.C. 621.396.946 : 629.19

Optimum System Engineering for Satellite Communication Links with Special Reference to the Choice of Modulation Method

By

W. L. WRIGHT, B.A.†

AND

S. A. W. JOLLIFFE†

Presented at the Convention on "Radio Techniques and Space Research" in Oxford on 5th–8th July 1961

Summary: Single sideband, wideband f.m. and pulse-code methods are considered. Pulse code system is shown to have advantages since it more nearly approaches the ideal expressed by the Hartley-Shannon law. Expected signal/noise ratio values in a practical multi-channel (600 channel) satellite relay system are discussed.

1. Introduction

Existing global radio communication is chiefly by means of narrow-band h.f. transmissions within a crowded spectrum where regimentation is only partly successful. The limited bandwidth and the vagaries of ionospheric reflection preclude its serious consideration as a future primary service capable of meeting the anticipated demand, and the need for an improved system is now generally accepted.

Even before the launching of the first satellite, the possibility of improved point-to-point communications by means of satellite repeaters was foreseen.[1] Such a project, although at present perhaps of limited capability, now appears feasible since the components required are within the foreseeable state of the art. Within a decade, the first practical outcome from the vast expenditure on space research will no doubt take the form of improved communication and navigation facilities.

The military authorities undoubtedly regard the use of satellites as an essential component of their global navigation and communication networks. The concept that survival may depend on a vital message being received at a remote point at a precise time will undoubtedly result in satellites being launched to back up the existing communication links.

Whereas in meeting military communication demands cost is of little importance, the system which carries future long-distance commercial business will be that which provides the required number of toll quality circuits at the most economic rate. Thus, in order to retain and expand their business, the commercial operators must examine the possibilities of satellite communication links and compare them with the probable developments of existing systems, e.g. solid lines. It is not easy at present, due to lack of facts, to arrive at the relative economic merits of the two systems. Nevertheless it seems certain that satellites will at least provide the means for augmenting existing high quality circuits.

At present there are severe limitations in the level of r.f. power it is possible to provide in the satellite. Thus, the conservation of signal strength and the minimizing of noise are system design factors which dominate the satellite communication problem. The high cost of satellite launchings, and the present impossibility of undertaking maintenance of its component parts, make equipment reliability the most critical system engineering factor, and is decisive in the economics sum.

World-wide real-time satellite communication systems must employ several satellites. It is an advantage if these satellites maintain the same positions relative to each other. It may also be an advantage if the satellite attitude is closely controlled. Measures of control of both parameters are possible,[2] but their reliability has yet to be proved. In order to secure a high information rate it will be necessary to use high gain aerials on some satellites. In such systems a close control of attitude and position stability will be highly desirable.

The information capacity of a radio link is determined by the information bandwidth and by the signal-to-noise ratio present at the output of the receiver. This latter ratio is the product of a number of factors including the equivalent noise temperature of the receiving system and the efficiency of the modulation method employed, but it is also a function of the transmitted power.

Due to the considerable cost per unit mass of launching, and limitations of existing boost vehicles, the present generation of satellites are of low weight. The available satellite weight must be divided between the structure, the attitude control system, the power supply, the communication equipment and possibly a temperature and station-keeping control system. Whilst it would appeaı desirable to apportion as much

† Marconi's Wireless Telegraph Co. Ltd., Research Laboratories, Great Baddow, Essex.

weight as possible to the communication equipment and power supplies, this is only possible up to a point where the auxiliary services allowance becomes inadequate to support the operational requirements. In time, vehicles capable of launching greater payloads will become available, resulting in systems having a greater information capacity.

The revenue earning potential of any commercial point-to-point communication system can be judged by its information rate capability. Communication systems are thus required to transmit at as high an information rate as possible within the limitations of a given transmitter power and system bandwidth. For reasons already stated, this is of extreme importance in the satellite communication case, especially in the satellite-to-ground path.

Since the information capacity of a communication system is greatly influenced by the method of modulation employed, the relative efficiencies of various possible modulation systems are examined in the sections which follow.

2. Efficiency of Communication Systems

Communication systems are required to transmit at as high an information rate as possible within the limitations of a given transmitted power and system bandwidth. This is of extreme importance in the communication satellite case, especially in the satellite-to-ground path.

Both the method of modulation and the type of message processing used determine the efficiency of any communication system. System efficiency may be defined in a number of ways,[3] but that which will be adopted for the present discussion will be based on the Hartley-Shannon[4, 5] "channel capacity" equation

$$H = B\log_2(1+S_B) \qquad \ldots\ldots(1)$$

which relates the quantities

H = information rate in bits per second;

B = bandwidth of the channel in cycles per second;

S_B = signal/noise ratio in bandwidth B

$\left(S_B = \dfrac{P_r}{KTB}\right.$ where P_r = received power

KT = 1·38 joules × receiving system effective noise temperature in degrees Kelvin).

Consider first a radio communication system transmitting in a radio frequency bandwidth B at such a power level that at the distant receiver input a signal/noise power ratio S_B exists.

According to eqn (1) the equivalent bit rate H_0 at this point in the system is $B\log_2(1+S_B)$. If information in bandwidth M is carried as a modulation of the r.f. waveform, the post detector output of the receiver will in general have an information rate $H_M = M\log_2(1+S_M)$ where S_M is the receiver output signal/noise ratio in bandwidth M.

In an ideal system of modulation and demodulation the information rates are the same at both input and output points of the receiver.

i.e. $$H_0 = H_M$$

or $$B\log_2(1+S_B) = M\log_2(1+S_M)$$

Therefore $$(1+S_M) = (1+S_B)^{\alpha} \qquad \ldots\ldots\ldots(2)$$

where $$\alpha = \frac{B}{M}$$

i.e. for an *ideal system*, post-detector

S/N (dB) $\simeq$ (pre-detector S/N) (dB) ....(3)

In the case of most systems, eqns (2) and (3) will not hold since in general H_M is less than H_0.

In any practical system, by raising the power of the distant transmitter from P_T to P'_T (and consequently increasing the received power from P_r to P'_r), the resulting new output information rate H'_M may be made to equal H_M, the information rate for transmitted power P_T in an ideal system.

The efficiency factor here suggested is based on the r.f. power required in a practical system in comparison with that required in an ideal system of equivalent output information capacity, i.e.

System efficiency factor β =

$$\frac{\text{transmitter power required for an ideal system}}{\text{transmitter power required in an actual system}} = \frac{P_r}{P'_r}$$

when the bit rate per cycle per second (i.e. H/B bits per cycle) is the same in each case and the receiver equivalent noise temperatures are also the same.

Equation (2) is rewritten for the actual or non-ideal case, as follows:

$$(1+S'_M) = (1+S_B)^{\alpha}$$

or $$S'_M = (1+S_B)^{\alpha}-1 \qquad \ldots\ldots(4)$$

where S'_M corresponds to the increased r.f. S/N ratio S'_B required in an equivalent non-ideal system.

In the receiver there will be some relationship between the S/N ratio of the modulated r.f. signal entering it (S_B) and that of the demodulated signal leaving it (S_M),

i.e. S_M is a function of S_B

which we will denote by $S_{M(S_B)}$ ......(5)

Therefore from (4),

$$S_{M(S'_B)} = (1+S_B)^{\alpha}-1 \qquad \ldots\ldots(6)$$

By definition,

$$\beta \equiv \frac{P_T}{P'_T} \equiv \frac{P_r}{P'_r} \equiv \frac{S_B}{S'_B}$$

and substituting S_B/β for S'_B in eqn. (6) we have

$$S_{M(S_B/\beta)} = (1+S_B)^{\alpha}-1 \qquad \ldots\ldots(7)$$

from which β may be derived, given the relationship between the pre-detector and post-detector signal/noise ratios in any system. It is seen that in general β is a function of S_B, the input signal/noise ratio, as well as being a function of the modulation/demodulation process.

3. Comparison of Various Modulation Systems

It is of now of interest to compare the efficiencies of various modulation systems in terms of their β values.

3.1. *Efficiency of Single Sideband Modulation*

In this case the r.f. bandwidth, $B =$ the message bandwidth, M.

$$\alpha = \frac{B}{M} = 1$$

output S/N = K(input S/N)

i.e. $$S'_M = K\,.\,S'_B$$

or $$S'_M = \frac{K\,.\,S_B}{\beta}$$

where K is the detector efficiency.

From eqn. (7)

$$\frac{K\,.\,S_B}{\beta} = (1+S_B)-1$$

$$= S_B$$

i.e. $$\beta = K.$$

As might be expected of a system using this simple type of modulation, which is basically a linear frequency shift of the modulation spectrum to the radio frequency band and then a shift back again to the modulation band, the efficiency as defined by β is equal to that of the detector in the receiver.

Since the factor K may be made to approach unity, even at low values of r.f. signal/noise ratio, in systems which employ multiplicative mixing of the incoming signal by a reconstituted carrier (e.g. homodyne, synchrodyne and orthogonal systems), a s.s.b. system may be regarded as having an efficiency $\beta = 1$ for all practical values of input S/N. This is shown in curve (a) of Fig. 1.

It should be noted that nothing has been said regarding the ability of the system to meet a particular output S/N requirement. The ability to exchange bandwidth for output S/N is not a property of a simple s.s.b. system. If the output S/N must be raised above that which appears at the receiver output in an s.s.b. system, then more complex modulation methods must be resorted to, e.g. f.m., p.m., p.c.m., p.p.m. or p.w.m. In these systems, however, some loss of efficiency occurs in the process of exchanging bandwidth for signal/noise, so that the communication efficiencies of such systems are not as high as that of the s.s.b. case.

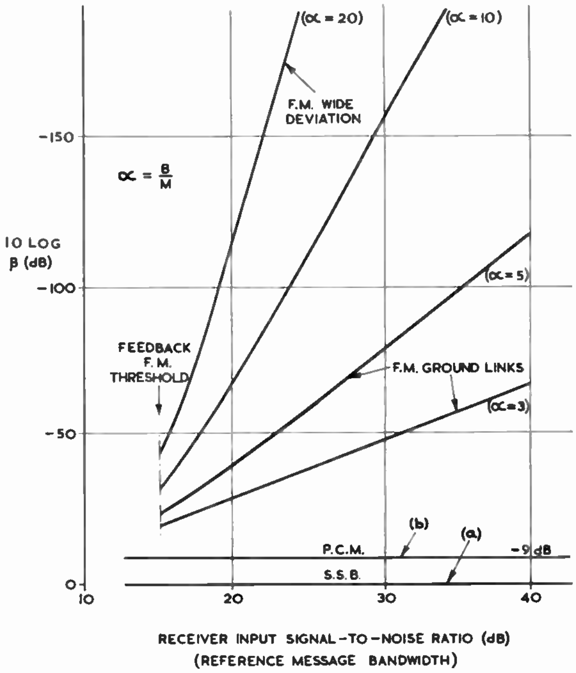


Fig. 1. Efficiencies of various modulation systems.

The next example, that of f.m., will illustrate this.

3.2. *Efficiency of Frequency Modulation*

The input/output S/N relation for f.m. is given by the well-known "f.m. improvement" equation:

$$S_M = \frac{3}{2} S_B \alpha \left(\frac{\Delta f}{f_m}\right)^2 \qquad \ldots\ldots(8)$$

where $\Delta f =$ peak deviation

$f_m =$ highest frequency in modulation band (0–M) c/s.

Consider the loss of communication efficiency incurred in S/N improvement.

In an f.m. system the minimum bandwidth B is given by

$$B = 2(\Delta f + f_m)$$

i.e. $$\frac{\Delta f}{f_m} = \left(\frac{B}{2f_m} - 1\right) = \left(\frac{\alpha}{2} - 1\right) \qquad \ldots\ldots(9)$$

Substituting the value of $\Delta f/f_m$ given by (9) into (8),

$$S_M = \frac{3}{2} S_B \alpha \left(\frac{\alpha}{2} - 1\right)^2 \qquad \ldots\ldots(10)$$

Equation (10) is thus the full expression of the function $S_{M(S_B)}$ for the case of an f.m. system.

Rewriting eqn (7) according to (10)

$$\frac{3\alpha}{2}\left(\frac{\alpha}{2} - 1\right)^2 \frac{S_B}{\beta} = (1+S_B)^\alpha - 1$$

Therefore

$$\beta = \frac{\frac{3\alpha}{2}\left(\frac{\alpha}{2} - 1\right)^2 S_B}{(1+S_B)^\alpha - 1} = \frac{\frac{3}{2}\left(\frac{\alpha}{2} - 1\right)^2 S_0}{\left(1 + \frac{S_0}{\alpha}\right)^\alpha - 1} \qquad \ldots\ldots(11)$$

where S_0 is the input S/N ratio referred to an r.f. bandwidth of M, the message bandwidth.

The efficiency, as might be expected, varies with α, the factor which also governs the exchange of bandwidth for output S/N. For a given value of α, β decreases as the input S/N is increased. This is shown in Fig. 1 in which plots of eqn (11) are made showing the variation of 10 log β against 10 log S_0 for various values of α. The larger the value of α, i.e. the greater the S/N improvement, the worse the efficiency becomes. (The f.m. threshold indicated in Fig. 1 is shown at 10 log S_0 = 15 dB since frequency following at the receiver is assumed in which B in the limit approximates to $2M$, at which bandwidth the r.f. S/N ratio is $S_0/2$ and this should correspond to 12 dB.)

For instance, at an input S/N ratio αS_B of 20 dB and with a value of $\alpha = 5$ (typical of point-to-point ground links), the S/N improvement over s.s.b., (S_M/S_0) is about 5·5 dB (from (10)), but the efficiency needed for this S/N improvement—which depends on the increase in bandwidth—is, from Fig. 1,

$$10 \log \beta = -40\,\text{dB}$$

or

$$\beta = 10^{-4}$$

The efficiency is even lower for wideband f.m. systems (as contemplated for use in many proposed satellite communication systems) in which α factors of 20 are typical. Here the f.m. improvement over s.s.b. amounts to 21 dB, and, working at an input S/N ratio of αS_B = 20 dB as before, we obtain a β value of −115 dB, i.e. the equivalent f.m. system would require a transmitter power level of 115 dB above that of the transmitter in an ideal system, working at the same information rate within the same radio frequency bandwidth.

Clearly a high and disproportionate price in bandwidth is paid in f.m. systems—and an especially high price in wide-deviation f.m. systems—for the increased output S/N ratio obtained.

3.3. *Efficiency of Pulse Code Systems*

Pulse code is not in itself a system of modulation; it is a particular form in which information is expressed. In discrete systems the information to be transmitted is generated in quantized form (e.g. teleprinter signals) and may be transmitted with or without further encoding. Information such as speech, music or television is usually conveyed by transmitting the voltage analogue of the information in a continuous system. If, however, this analogue information is sampled, usually at a rate equal to twice the highest component of the signal spectrum, quantized by passing it through an amplitude gate which recognizes the sample as one of L discrete levels, then $\log_2 L$ bits per sample are produced and the information appears as pulses of L different amplitudes.

Each of these L level samples may be represented by a group of n pulses, each of which has b possible amplitudes such that $L = b^n$. If $b = 2$, each sample is denoted by a group of n binary digits. In this way analogue information may be changed to digital form. The chief purpose in effecting this change is to render the signal more discernible from noise.

For example, if it is required to transmit an analogue waveform with distortion not in excess of −40 dB, one must reproduce the waveform with an accuracy of 1%, necessitating approximately 100 quantizing levels or 7 binary digits per sample to convey the information with sufficient accuracy.

In the presence of noise and interference it is possible to recognize the binary code with far less probability of error than that involved in determining the instantaneous amplitude of the analogue signal to an accuracy of 1%. Provided that the receiver recognizes the code correctly in the presence of a given noise level, then the information may be reproduced at that point having suffered no ultimate degradation due to noise in the transmission system. Moreover, the pulses may be regenerated for further transmission so that the noise and interference present on any transmission path is not added in full to that on the succeeding path.

It is assumed that a pulse code waveform is made to shift the phase of the transmitter carrier through 180 deg each time the binary code changes from 0 to 1 or vice versa.

The S/N ratio of the pulse code waveform at the detector output in bandwidth nM, after demodulation by a phase detector supplied with a synchronous reference carrier, is then about 3 dB above the r.f. S/N ratio at the receiver input.

The equivalent bit rate for pulse code is given in eqn. (4) of Reference 6 by the approximate relationship,

$$H \simeq B \log_2 \left(1 + \frac{S_B}{8}\right) \qquad \text{......(12)}$$

For an ideal system $H_0 = B \log_2 (1+S_B)$. If H' is made to equal H_0 by increasing S_B in eqn (12) to S'_B, then

$$\frac{S'_B}{8} \simeq S_B$$

or

$$\beta \equiv \frac{S_B}{S'_B} \simeq \tfrac{1}{8}$$

β for pulse code systems is, according to eqn. (12), constant for all input S/N ratios and represents about 9 dB loss in signal power in comparison with the ideal system. (See Fig. 1, curve (b).)

Further consideration will now be given to pulse code modulation in respect of the output signal/noise ratio obtained with a given r.f. input signal/noise ratio.

4. S/N Ratio Obtained at the Output of a P.C.M. System

In a *binary p.c.m. system* it can be shown[7] that the signal/noise (P) at the output of the decoder, due to code *error probability* in the presence of noise, is given by

$$P = \frac{3\sqrt{3}\, m^2}{8 \;\; f_0\tau} \, . \exp \tfrac{1}{2}\left(\frac{V}{e}\right)^2 \qquad \text{......(13)}$$

$$= 0{\cdot}65 \frac{m^2}{f_0\tau} \, . \exp \tfrac{1}{2}\left(\frac{V}{e}\right)^2$$

where $2V$ = peak-to-peak amplitude of the code pulses as they enter the decoder

e = r.m.s. value of the noise voltage at the decoder input

f_0 = cut-off frequency of the ideal low pass filter preceding the decoder

τ = time occupied by one pulse

m = degree of modulation.

In the above equation a binary code and a sinusoidal modulating waveform are assumed.

Assuming also that $m = 1$ and $f_0\tau = 0{\cdot}5$ (the acceptable minimum value, this factor determining the shape and amplitude of a pulse at the output of the filter), then from (13) the following expression is obtained

$$10 \log P \simeq 10 \log 1{\cdot}3 + \tfrac{1}{2}\left(\frac{V}{e}\right)^2 . \, 10 \log \varepsilon$$

$$\simeq 1{\cdot}1 + 2{\cdot}17\left(\frac{V}{e}\right)^2$$

$$\simeq 2{\cdot}2\left(\frac{V}{e}\right)^2 \qquad \text{......(14)}$$

If the binary code is transmitted by a 180 deg phase shift of a radio frequency carrier, then

$$\left(\frac{V}{e}\right)^2 = 2S_B$$

where S_B is the r.m.s. carrier/noise power in an r.f. bandwidth B equal to f_0

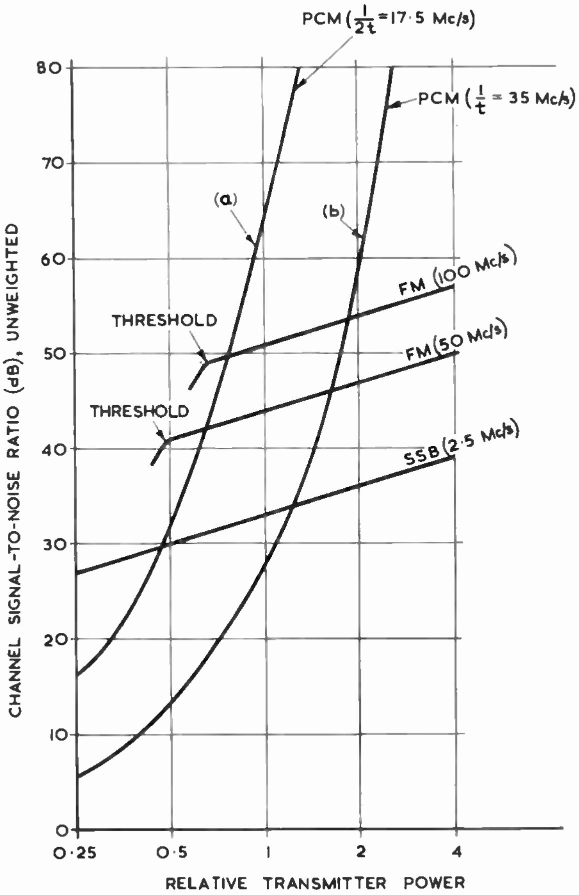
From (14)

$$10 \log P \simeq 4{\cdot}4 S_B \qquad \text{......(15)}$$

It can be shown that the minimum bandwidth of a p.c.m. system is given by multiplying the highest modulation frequency by the number of pulses required to code each sample,

i.e.

$$B = nM \qquad \text{......(16)}$$



NOTE:- THE POWER SCALE IS IN WATTS FOR THE SYSTEM TAKEN IN THE TEXT AS AN EXAMPLE.

Fig. 2. Channel output signal/noise ratios for three types of modulation.

Equation (15) may therefore be written as

$$10 \log P \simeq \frac{4 \cdot 4}{n} S_0 \qquad \ldots\ldots(17)$$

S_0 is the S/N ratio which would be obtained at the receiver input if measured in the message bandwidth M, i.e. it is that which is applicable to an s.s.b. system.

The second source of noise which appears at the decoder output is due to quantization. If Q is the r.m.s. signal output power to r.m.s. quantization noise output power ratio, then it can be shown that

$$Q = \tfrac{3}{2}4^n \qquad \ldots\ldots(18)$$

or

$$10 \log Q = 1 \cdot 8 + 6n \qquad \ldots\ldots(19)$$

4.1. *Equalization of Thermal and Quantization Noise*

It is desirable in practice to arrange n so that thermal and quantization noise contributions are approximately equal at the expected normal signal input level, less a prescribed fading margin, i.e. $P = Q$ at an equivalent s.s.b. input S/N ratio S_0 where S_0 represents the ratio after allowing for the fading margin.

When $P = Q$

$$P = \frac{3\sqrt{3}}{4} \exp\left(\frac{S_0}{n}\right) = \tfrac{3}{2}4^n \qquad \ldots\ldots (20)$$

from (13) and (18)

or

$$\frac{\sqrt{3}}{2} \exp\left(\frac{S_0}{n}\right) = 4^n$$

Therefore

$$n^2 \simeq 0 \cdot 72 S_0$$

or

$$n \simeq 0 \cdot 85(S_0)^{\frac{1}{2}} \qquad \ldots\ldots(21)$$

Substituting this expression for n into eqn (18),

$$P = Q = \tfrac{3}{2} \times 4^{0 \cdot 85(S_0)^{\frac{1}{2}}}$$

or

$$10 \log P = 10 \log Q \simeq 1 \cdot 8 + 5(S_0)^{\frac{1}{2}} \qquad \ldots\ldots(22)$$

The logarithmic relationship between S_0 and P is shown in curve (*a*) of Fig. 2, which is a plot of eqn (22), S_0 being shown in terms of corresponding transmitter power.

In the section which follows examples are given in which calculations of required minimum r.f. powers and bandwidth occupancies are made for the three types of modulation system whose efficiencies have been compared above.

5. Expected Signal/Noise Ratios in Practical Multi-channel Satellite Relay Systems

5.1. *System Parameters*

An example is taken from an American proposal made in July 1960 for a wide-deviation *f.m. system.*[8] The system geometry is as follows:

Satellite in circular polar orbit, not stabilized in attitude:	Height 2500 miles.
Minimum path distance:	2500 miles.
Maximum path distance:	4600 miles.

The signal/noise ratio in the highest channel of a 600-channel f.d.m. system is derived below, assuming the following parameters:

Radio frequency:	6000 Mc/s.
Satellite transmitter power:	1W.
Satellite aerial gain:	−3 dB.
Ground receiver aerial diameter:	60 ft. (Gain 58·8 dB assuming 0·6 efficiency factor.)
Noise temperature of ground receiving system:	30° K.

(Galactic noise is negligibly small at 6000 Mc/s. Earth's atmosphere and maser noise constitute the chief contribution to this noise temperature.)

A frequency-following receiver is used on the ground having an i.f. bandwidth somewhat greater than twice the highest modulation frequency, i.e.

$$\begin{aligned} B_{IF} &= 2 \times 2 \cdot 54 \times \text{bandwidth margin factor (say } 1 \cdot 5) \\ &= 7 \cdot 62 \text{ Mc/s} \\ &\simeq 8 \text{ Mc/s.} \end{aligned}$$

5.2. *Satellite-to-Ground Path*

Calculation of signal/noise ratio at input of receiver limiters in bandwidth, $B_{IF} = 8$ Mc/s.

Path attenuation $\alpha = 96 \cdot 6 + 20 \log (DF)$

where D is distance in miles

F = frequency in kMc/s

Taking D as the maximum distance, 4600 miles,

$$\begin{aligned} \alpha &= 96 \cdot 6 + 20 \log (4600 \times 6) \\ &= 96 \cdot 6 + 88 \cdot 8 \\ &= 185 \cdot 4 \text{ dB} \end{aligned}$$

Aerial gains $= 58 \cdot 8 - 3$

$= 55 \cdot 8$

Hence, with 1W transmitted power, received power level is

$- 129 \cdot 6$ dBW

say $- 130$ dBW

Noise power in 1 Mc/s bandwidth at 30° K

$= -154$ dBW

Therefore S/N (maximum distance)

= 24 dB in 1 Mc/s bandwidth

or 48 dB in 4 kc/s bandwidth

Margin above threshold (*f.m.*)

S/N in i.f. bandwidth

(8 Mc/s) = 24 dB − 9 dB

= 15 dB

The threshold level for f.m. working is about 12 dB so that the working margin above threshold is +3 dB, showing that with these parameters proper f.m. operation is possible provided that a frequency-following receiver is employed in receiving the signals from the satellite.

5.3. 600 *Channel System: Channel Deviation*

Allowing a bandwidth margin factor of 1·5 and using the figure of 2·4 for the crest factor of the multi-channel signal, the permissible r.m.s. multi-channel deviation, $\Delta f = 13$ Mc/s (peak value 31 Mc/s) when the total r.f. bandwidth occupancy is 100 Mc/s.

The corresponding channel r.m.s. deviation due to a 1 mW test tone applied at a point of zero reference level, δf, is 15 dB below this, or $\delta f = 2{\cdot}3$ Mc/s.

Substituting this value of δf in the f.m. equation,

$$\frac{\text{r.m.s. signal power in test tone of 1 mW}}{\text{noise power in the top 4 kc/s channel}} \equiv \frac{P_s}{P_n}$$

$$= \frac{\text{received r.f. signal power}}{\text{input noise in 4 kc/s bandwidth}} \times \left(\frac{\text{r.m.s. deviation due to 1 mW test tone}}{\text{channel centre frequency}}\right)^2$$

we obtain,

$$10 \log \frac{P_s}{P_n} = 48 + 20 \log \frac{2{\cdot}3}{2{\cdot}54}$$

$$= 47 \text{ dB for the highest (and worst) channel.}$$

If pre-emphasis is applied to even out more nearly the S/N in the several channels, whilst maintaining the same multi-channel deviation, an extra 4 dB is added to the S/N figure in the top channel.

i.e. $$\frac{P_s}{P_n}\text{(with pre-emphasis)} = 51 \text{ dB.}$$

A level of 46 dBm0 measured unweighted in a 4 kc/s telephone channel, corresponding to 49·5 dBm0 psophometrically weighted in a 3·1 kc/s channel, should result in an acceptable overall channel S/N ratio for a complete ground-to-satellite-to-ground link. The figure of 51 dBm0 unweighted, obtained above for an f.m. system, gives a 5 dB margin above this level.

The large bandwidth requirement of 100 Mc/s per one-way path of a 600-channel satellite f.m. system makes it unlikely that a sufficiently wide band of frequencies will be allocated to accommodate four such bands in order to provide two-way working on both ground-to-satellite and satellite-to-ground paths. When guard bands are accommodated, the total frequency block would amount to about 500 Mc/s.

It is possible that a compromise might be made, in which a total band of about 200 Mc/s might be allocated. The deviation corresponding to such a system would result in a channel signal/noise ratio of about 43 dB unweighted, compared with the 51 dB calculated for the wider bandwidth system.

It should be noted that frequency-following has been assumed in the ground station receivers. This is necessary in order to preserve a received signal/noise ratio in excess of the f.m. threshold. While such techniques have been used successfully in systems having modulation bandwidths up to about 0·5 Mc/s, further development will be required in order to extend this so that the feedback loop will operate with stability over a 5 Mc/s modulation band appropriate to a television channel or even over the minimum bandwidth of 2·5 Mc/s required for a 600 telephone channel link.

5.4. *S/N Using Other Systems of Modulation*

5.4.1. Single sideband

Keeping all other parameters of the system the same as assumed above, but changing over to s.s.b. modulation (and allowing a 15 dB figure for the multi-channel signal power to the single channel power ratio as before), the s.s.b. 600-channel figures become:

Satellite transmitter mean power:	1W
Total bandwidth:	2·5 Mc/s
Channel S/N:	33 dB

These figures assume that the transmitter is capable of radiating a peak power of 6W during peaks of the multi-channel signal, of which the crest factor (exceeded for 1% of the time) = 2·4 (i.e. +7·6 dB).

If 46 dB S/N is required, then

transmitter mean power = +13 dBW = 20W
and transmitter peak power = 115W.

5.4.2. Pulse code modulation

Figures 3, 4 and 5 refer to p.c.m., illustrating different ways of sampling the 600 channels and combining them into a pulse-coded signal.

Figure 3 shows the standard channelling equipment to develop the 600-channel f.d.m. signal in a 2·5 Mc/s baseband. This is sampled at a 5 Mc/s rate and each sample is encoded to form a 7-binary digit train. The pulse rate is therefore 35 megabits per second. This may be transmitted in a *minimum* bandwidth of 17·5 Mc/s. It should be noted that the number of digits per coded sample has been adjusted according to the equations given in the section above on S/N in p.c.m. systems, to give approximately equal output contributions from thermal noise and quantization noise.

For 1W transmitted power from the satellite, the S/N in 1 Mc/s bandwidth = 24 dB (as calculated above). Therefore S/N in 17·5 Mc/s bandwidth = 11·5 dB.

Allowing a fading margin of 1·5 dB brings this figure to 10 dB.

5.5. *Output S/N (Thermal)*

By using carrier phase reversal to distinguish between zeros and ones, a further 3 dB increase in S/N may be achieved after demodulation by a phase detector. (Phase shift keyed (p.s.k.) system.) Therefore S/N after demodulation but before decoding = 13 dB.

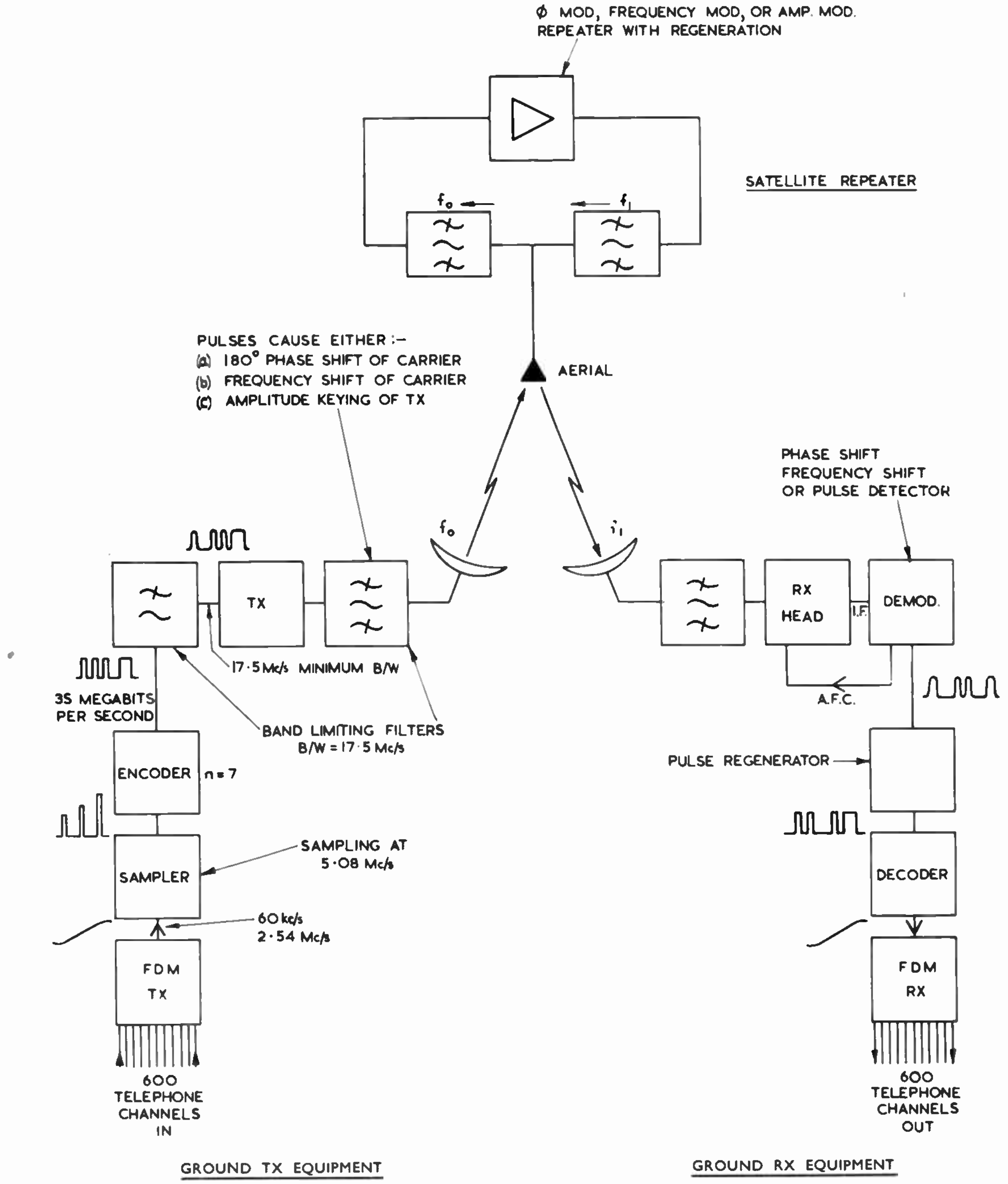
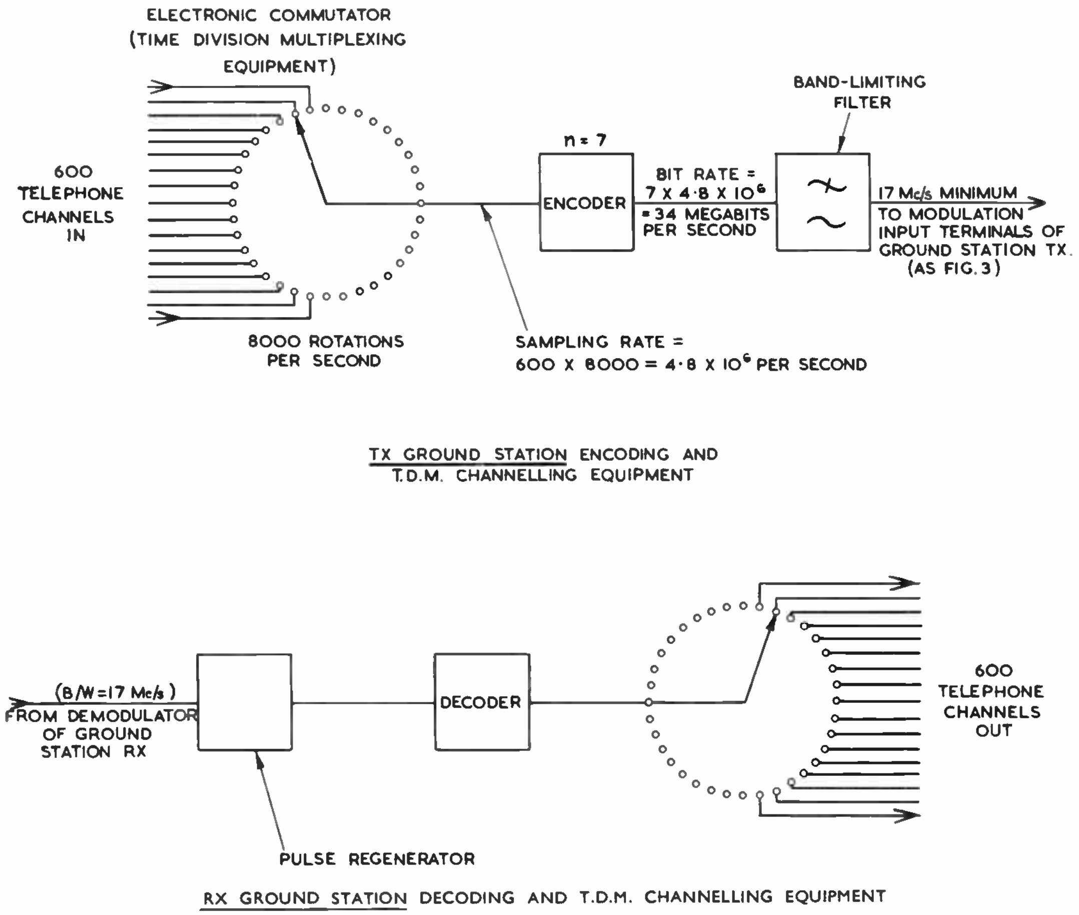


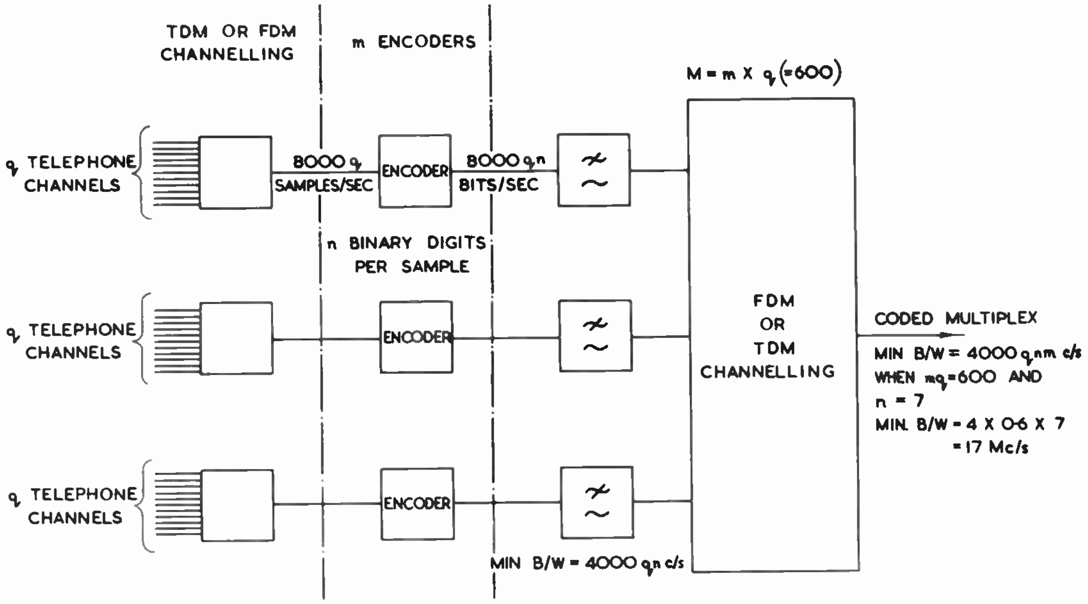
Fig. 3. 600-channel p.c.m. communication system, employing frequency division multiplexing before encoding.



TX GROUND STATION ENCODING AND T.D.M. CHANNELLING EQUIPMENT

RX GROUND STATION DECODING AND T.D.M. CHANNELLING EQUIPMENT

Fig. 4. 600-channel p.c.m. communication system. Preferred method to that of Fig. 3, employing time division multiplexing before encoding.



Note: Figs. 3 and 4 represent limiting cases when m = 1 and q = 600; the opposite extreme case is when m = 600 and q = 1

Fig. 5. Multiplexing and encoding of *M* telephone channels: general case for p.c.m.

Applying eqn (14) we obtain for the S/N ratio after decoding,

$$10 \log P = 1{\cdot}1 + (2{\cdot}17 \times 20)$$
$$= 44{\cdot}5 \text{ dB.}$$

5.6. *Output S/N (Quantization)*

Peak-to-peak signal power/r.m.s. quantization noise power = $10{\cdot}8 + 6n$ dB

where n is the number of binary digits

$$= 10{\cdot}8 + (6 \times 7) = 52{\cdot}8 \text{ dB.}$$

Converting peak-to-peak signal power to r.m.s.

$$\frac{\text{Signal power}}{\text{Noise power}} = 52{\cdot}8 - 9 \text{ dB}$$
$$= 44 \text{ dB.}$$

When using the above method of sampling, this quantization noise would result in inter-channel modulation of approximately equivalent level.

The parameters then for 600-channel p.c.m. are:

Satellite transmitter power	= 1W
Bandwidth (minimum)	= 17·5 Mc/s
S/N (thermal)	44 dB
S/N (quantization)	44 dB.

Exactly the same figures are obtained for satellite transmitter power, S/N and bandwidth if we use other methods of combining the 600 channels as shown in Figs. 4 and 5.

6. Conclusions

(1) A satellite transmitter power of between 1 and 5 watts should be sufficient for 600 telephone channel communication or a single television channel.

The ground-to-satellite transmission path is not considered to present difficulty, since the ground transmitter may have a large output power.

(2) The selection of the most suitable modulating system is very important in satellite communication systems.

(3) Single sideband seems an attractive system from the bandwidth and efficiency points of view, but the peak power demand and the very tight specification placed on the amplitude linearity of the satellite repeater are features which weigh heavily against it.

(4) Frequency modulation, as far as ground links are concerned, is a well-tried modulation system for multi-channel telephony and television purposes.

F.m. has the advantage of employing apparatus in the satellite and on the ground which is not unduly complicated and the design parameters of which are well understood—with the following exception:

In the case of the ground receiver, the feedback technique for phase lock or frequency following over a wide modulation bandwidth (2·5 Mc/s for multi-channel or 5 Mc/s for television), required to maintain the received signal above the 12 dB f.m. threshold, is not easy and further development work will be necessary to maintain stability over this wide bandwidth.

Wide-deviation f.m., however, such as would be required for transmitter powers in the range 1 to 5 watts would occupy what appears to be a prohibitive bandwidth and makes very inefficient use of the 100 Mc/s or so channel frequency allowance in comparison with the theoretical information capacity appropriate to such a bandwidth.

(5) P.c.m. is a powerful aid in helping to achieve the high output S/N ratio, required of an inter-continental communications link, when the r.f. conditions are such that the S/N ratio in an equivalent s.s.b. system would be about 20 dB only.

From the equations given it is seen that the output S/N (thermal) ratio and the output S/N (quantization) ratio are each above 51 dB under such conditions.

The assumption that a bandwidth of $1/2t$ (where t = pulse width) may be used in which to transmit the pulse information whilst still retaining the required low level of inter-pulse cross-talk information, constitutes one of the more dubious points in the discussion, but it must be understood that this bandwidth is intended to represent a *target* value which might possibly be achieved by refined methods.

If in practice the more readily achievable bandwidth of $1/t$ were used, then a 3 dB increase in transmitter power and a corresponding increase in bandwidth (bandwidth = 35 Mc/s in the example given instead of 17·5 Mc/s) would result in a S/N output ratio 3 dB lower than that which could have been achieved with a bandwidth of $1/2t$ at normal power level.

It is unnecessary to elaborate on the well-known advantages of p.c.m. arising out of pulse regeneration at the repeaters and receiving terminal, except to note that by this means cumulative noise and distortion in an n-path system are avoided. Equipment advantages in the repeater (the satellite) are similar to those applicable to the f.m. case in which amplitude limiting may take place with advantage.

Experiments with a recently designed 60-channel p.c.m. equipment are in progress with a view to further assessment of pulse code techniques in practical systems, including scatter link circuits having intrinsically low S/N ratios. It is hoped that confirmatory results will be obtained to strengthen what appears to be a very good case for the employment of p.c.m. in active satellite communication systems.

In this paper factors have been discussed which determine the system capacity of communication links employing earth satellite repeaters. The most that can be said at present is that such systems are theoretically possible. Useful satellite systems will be technically feasible when it has been demonstrated that the equipment will work for long periods in the space environment. If it can, in the course of time, be shown that the assured life of a satellite repeater is of sufficient duration to convey information more economically than by other means, this new mode may eventually monopolise world commercial communication traffic.

7. Acknowledgments

The authors gratefully acknowledge the help of their colleagues who are engaged in work concerning the use of radio in space activities. The paper is published by permission of the Director of Research, Marconi's Wireless Telegraph Co. Ltd.

8. References

1. A. C. Clarke, "Extra-terrestrial relays", *Wireless World*, **57**, pp. 305–8, October 1945.
2. W. F. Hilton and B. Stewart, "The advantages of attitude stabilization and station keeping in communications satellite orbits", *J. Brit.I.R.E.*, **22**, pp. 193–202, September 1961.
3. R. W. Saunders, "Communication efficiency comparison of several communication systems", *Proc. Inst. Radio Engrs*, **48**, pp. 575–88, April 1960.
   A. J. Viterbi, "Coded phase-coherent communications", *Trans. Inst. Radio Engrs* (*Space Electronics and Telemetry*), **SET-7**, pp. 13–4, March 1961.
4. R. C. L. Hartley, "Transmission of information", *Bell Syst. Tech. J.*, **7**, p. 535, July 1928.
5. C. E. Shannon, "A mathematical theory of communication", *Bell Syst. Tech. J.*, **27**, pp. 379–423 and pp. 623–656, July and October 1948.
6. B. M. Oliver, J. R. Pierce and C. E. Shannon, "The philosophy of p.c.m.", *Proc. Inst. Radio Engrs*, **36**, pp. 1324–31, November 1948.
7. A. G. Clavier, P. F. Panter and W. Dite, "Signal-to-noise improvement in a p.c.m. system", *Proc. Inst. Radio Engrs*, **37**, pp. 355–9, April 1949.
8. J. B. Fisk, J. R. Pierce, C. M. Mapes and B. McMillan, "Frequency Needs for Space Communication", A.T. & T. Co. F.C.C. Docket No. 11866, dated July 6th 1960,

Manuscript first received by the Institution on 1st July 1961 and in revised form on 10th February 1962. (Paper No. 728.)

OBITUARY

The Council has learned with regret of the deaths of the following members:

Alexander Victor Simpson who died on 17th January 1962 was one of the earliest Members of the Institution, having been elected in May 1926.

For much of his professional life Mr. Simpson was a lecturer in radio subjects at Bolton Technical College. During the war years he was engaged in research and experimental work on radio in the Department of Electro-Technics of the University of Manchester for the Ministry of Supply.

Mr. Simpson took an active part in furthering Institution affairs in the Manchester area. During the nineteen-thirties he acted as the local representative of the Institution and then as Honorary Secretary of the North Western Section from its foundation in 1940 until 1942.

Although in poor health for some years, Mr. Simpson continued his teaching until three months before his death at the age of 61 years. He leaves a widow.

* * *

John Edgar Gater-Jones was born at Swindon on 24th June 1896. During the First World War he joined the Royal Artillery and later transferred to the Signals Section of the Royal Engineers. From 1918–21 he served in the Royal Corps of Signals.

After demobilization he worked with various organizations as a radio engineer, and during this period he took an Advanced Communications Course at Coventry Technical College. In 1930 he joined the Ministry of Supply as an inspector of communications equipment; three years ago he became an Electronic and Radar Inspector in the Inspectorate of Electrical and Mechanical Engineering attached to A. C. Cossor. He was elected an Associate Member of the Institution in 1942. Mr. Gater-Jones died on 15th September 1961 after a brief illness.

* * *

Eric Boyle Mason died on 9th January 1962 in Manchester Royal Infirmary following an illness of some months duration. Mr. Mason had been with A. V. Roe and Company at Chadderton since 1957 as a radio designer. He was 39 years of age.

Mr. Mason received his technical training in the Royal Air Force and at the Polytechnic, Regent Street, London and following demobilization he was for four years with Philips Electrical. He then went to the United States for some three years where he was employed as an electronic engineer with Thomas Electronics of Passaic, New Jersey, on the manufacture and testing of cathode-ray tubes.

Registered as a Student of the Institution in 1944, Mr. Mason was transferred to Associateship in 1957.

Ernest Richard Wilcox died on 19th February 1962 as a result of a road accident. In 1939 he joined the R.A.O.C. as an armament artificer and was commissioned in the R.E.M.E. in 1941. He left the Army in 1961 with the rank of Major after holding various technical and staff appointments.

Major Wilcox was one of the R.E.M.E. representatives on the joint Brit.I.R.E.–R.E.M.E. Liaison Committee which was set up by the Council during the period 1960–61 to relate Institution qualifications and Corps courses.

In April 1961 Major Wilcox joined Wayne Kerr Laboratories as a research and development liaison engineer.

During the last five years of his army service he attended part-time courses at Hendon Technical College and subsequently obtained the Higher National Certificate with endorsements, thereby satisfying the Institution's examination requirements. Major Wilcox was elected an Associate Member in January 1961. He was 45 years of age and leaves a widow and two sons.

* * *

Donald Geoffrey Morley Alexander was born in Wandsworth, London, on 22nd July 1924. He was elected an Associate of the Institution in 1961.

During the war he served in the R.A.F. as a radar mechanic. In 1945 he joined Telephone Rentals and after nine years moved to English Electric Aviation as a junior engineer, concerned with the development of test gear and control equipment for guided weapons. In 1960 he joined the Plessey Company at Havant as a section leader in the Mobile Communications Group.

Since the war he had suffered from asthma and this contributed to his death on 26th December 1961 at the age of 37.

Mr. Alexander leaves a widow and a daughter.

* * *

Neil Bertram Acred, who was born on 12th December 1927 in Nottingham, received his technical education at the Huntingdon Evening Institute and the Nottingham and District Technical College. He obtained various City and Guilds Telecommunications certificates and was successful in completing the Institution's Examination, being elected a Graduate in 1951.

In June 1949 he joined Ericsson Telephones Limited as a laboratory design engineer, subsequently transferring to Bendix-Ericsson (U.K.) Limited with whom he remained until his death on 16th September 1961. He leaves a widow, two daughters and a son.

U.D.C. 621.396.933: 629.19

Navigation Satellites with particular reference to Radio Observations

By

W. A. JOHNSON, M.A.†

Presented at the Convention on "Radio Techniques and Space Research" in Oxford on 5th–8th July 1961.

Summary: The paper describes some work on radio observations on artificial satellites and the application to navigation problems in which the Royal Aircraft Establishment has participated.

1. Introduction

There is no novelty in the basic idea of using artificial satellites as navigational aids, by observing their orbits by analogy with lunar observations. We can, however, have multiple artificial satellites in selected orbits. Radio engineers can contribute observational methods applicable to artificial satellites which are independent of weather conditions, an important consideration in any navigation system. These observations can be used, in conjunction with highly accurate optical methods when conditions are favourable, for the determination of the satellite orbit in the first place, and subsequently for terrestrial position finding. This paper describes work in which the Royal Aircraft Establishment has collaborated.

2. Principles of Position Fixing

If the position of the satellite in space is known, all that is necessary is to determine its direction from the observer at that moment in order to be able to calculate position. The essential simplicity of this principle was demonstrated by W. A. Scott,[1] who observed the time at which the *Echo 1* satellite passed close to known stars. The complication arises, particularly in the subsequent computation, in that the predictions of the orbit were not sufficiently accurate for his purpose, and so he used observations on one transit to correct the orbit from his known position, and observations on the next transit to re-calculate his apparent position. The consistency obtained was within 1 mile in his position.

As Scott points out, the accuracy is limited principally by the lack of knowledge of the orbit, and present knowledge of perturbing effects does not enable predictions to be made far ahead. In particular, the fact that the earth is not a perfect sphere produces wiggles in the orbit, which can provide information of interest on the earth's equatorial bulge, but which must be compensated for in tracking a satellite for accurate navigation.

3. Radio Observations on Satellites Transmissions

At the Royal Aircraft Establishment we started making radio observations on transmissions from satellites as soon as the successful launching into orbit of *Sputnik 1* was announced. Particular attention was paid to accurate direction finding and observations

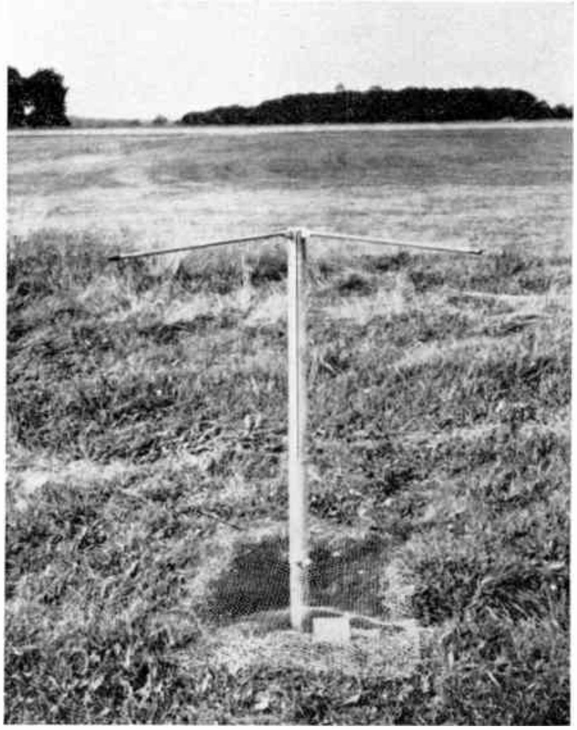


Fig. 1. A typical interferometer receiving aerial. This one is for 108 Mc/s band, and consists of two horizontal arms perpendicular to each other with a balanced feed. The aerial elements are somewhat over a quarter wave above the ground.

† Royal Aircraft Establishment, Radio Department, Farnborough, Hants.

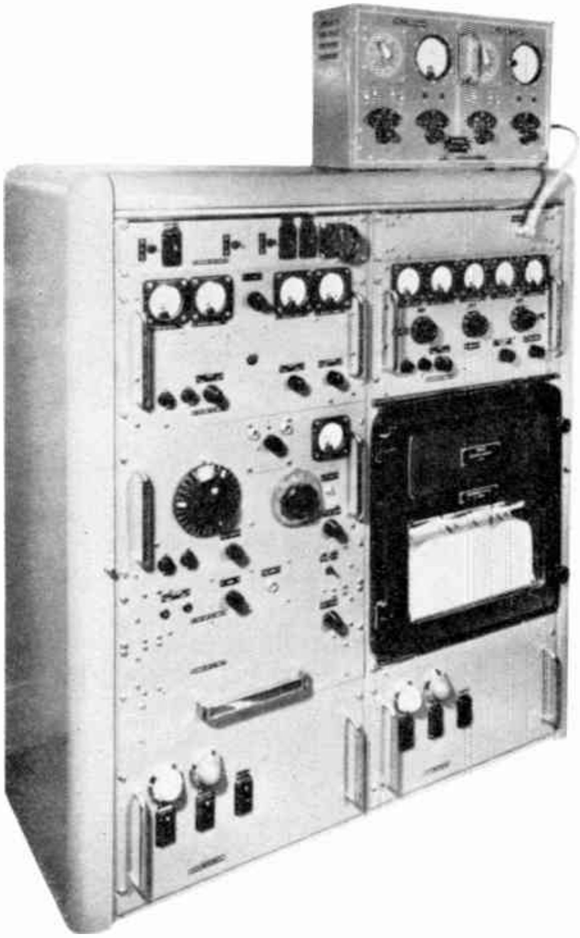


Fig. 2. Typical interferometer equipment. The instrument compares the phase of the signals received on pairs of aerials disposed East-West, and North-South and records the phase differences for subsequent analysis.

of the frequency of the received signal and these observations contributed to early orbit calculations.[2]

These two methods have been developed continuously since then and provide two complementary systems. The direction finding equipment used are now known as "interferometers", and the frequency observation techniques are referred to as Doppler methods.

3.1. *Interferometers*

If two widely spaced aerials are connected together by a suitable radio frequency transmission line, a multiple-lobe directivity diagram will result. In the receiving case, the signals from each aerial are superimposed on the transmission line, producing a standing wave or interference pattern. An interferometer measures this interference pattern, the position of nodes along the cable being related to the direction of arrival of the incident waves.

A spacing of many wavelengths is necessary in order to get high measurement accuracy. The multiple-lobe character of the directivity diagram however gives ambiguities in direction which are resolved in practice by using several pairs of aerials. A complete interferometer system consists of sets of aerials on two base lines, generally North/South and East/West, together with the equipment necessary to find the arrival direction of the signal.[2] Figure 1 is a single aerial element for a 108 Mc/s interferometer. Figure 2 shows one form of the equipment in which the information is recorded for subsequent analysis. It is quite possible to give an immediate visual display of the satellite bearing and Fig. 3 shows such a display unit.

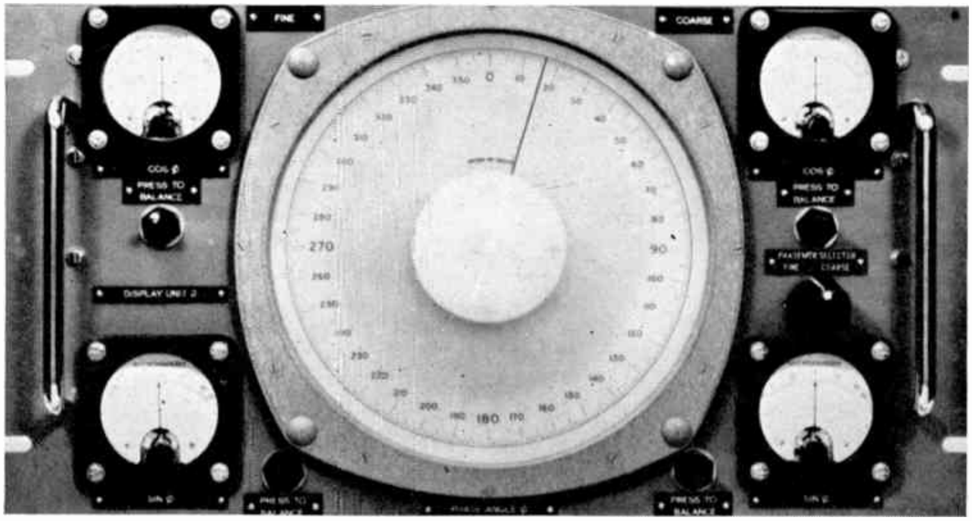


Fig. 3. A subsidiary display to the interferometer designed to give an approximate azimuthal bearing of the satellite in degrees.

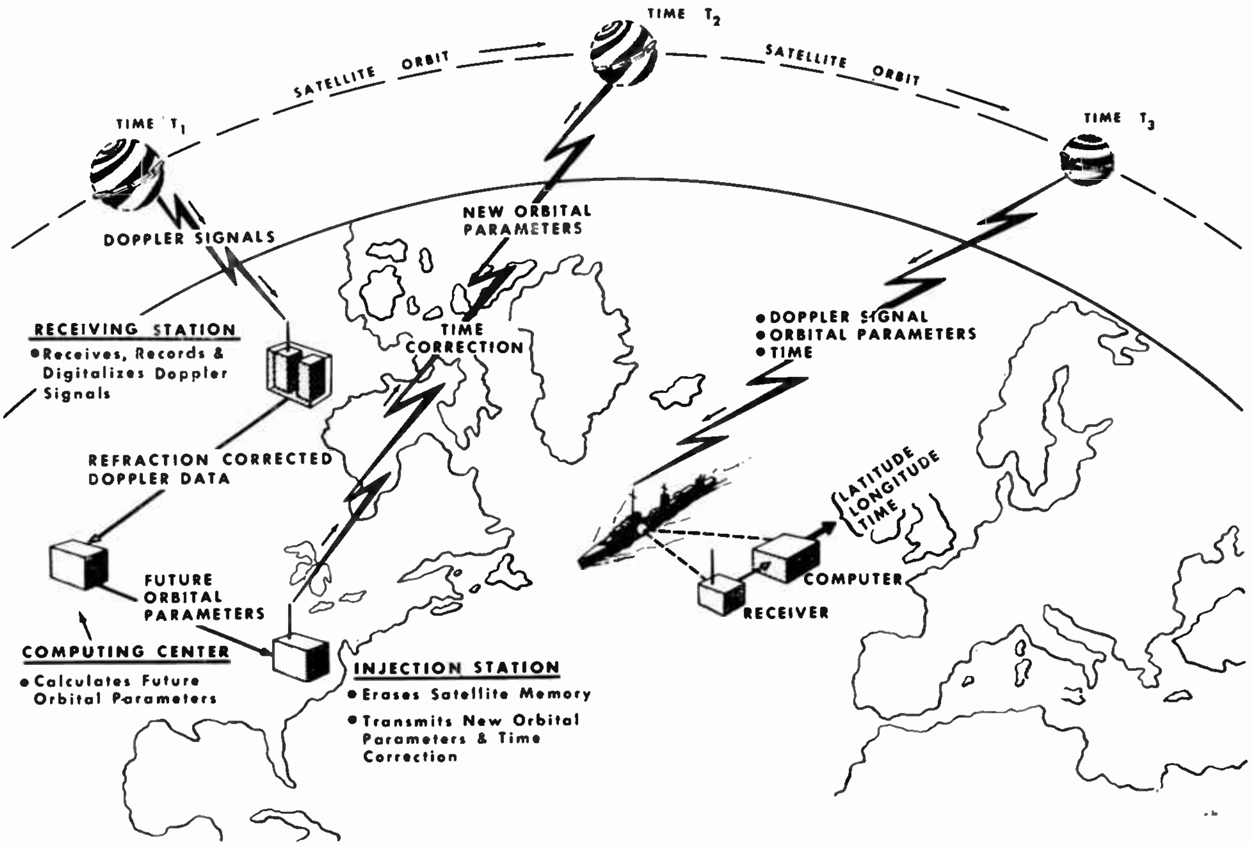


Fig. 4. Diagram showing the scheme of the *Transit* Project

(By courtesy of the Applied Physics Laboratory, Johns Hopkins University)

3.2. *Doppler Frequency Measurements*

At the same time as the early interferometers were being used, measurements on the frequency of the received signals from satellites were made. These when plotted against time give a Doppler curve, on which to first approximation the point of inflection might give the time of closest approach. Independently in the U.S.A., starting in late 1957 W. M. Guier and G. C. Weiffenbach explored the possibilities of using Doppler curves, and estimated in their paper[3] published in 1960 that a satellite Doppler system can provide navigation to at least one-half mile accuracy from a single Doppler curve.

These ideas were put to practical tests in a satellite programme in the U.S.A., which is leading towards a navigational system known as *Transit*. The Royal Aircraft Establishment has been co-operating since the summer of 1959 by taking measurements of Doppler curves on transmissions from test satellites.

4. Project "Transit"

This project is the satellite Doppler navigation system now being developed by the U.S.A. Navy Bureau of Weapons under contract with the Applied Physics Laboratory, Johns Hopkins University.[4] With only four satellites in orbit at any one time it appears possible to guarantee navigational fixes on nearly any part of the globe at least once every hour and a half. In the *Transit* operational system the satellites will orbit at altitudes optimum for accurate tracking, and tracking stations will receive the Doppler shift signals from each satellite pass within range and transmit these data to the computing centre.

The computing centre will calculate orbital parameters of the satellite for a minimum of 1 day ahead, and these data will be injected into the satellite for re-transmission by it once a minute. The navigational equipment will receive and record the Doppler shift of the satellite transmissions and the orbital data transmitted from the satellite, which will also transmit accurate time signals. From these data the navigator's latitude and longitude may be computed. Figure 4 shows the system operation schematically.

The Royal Aircraft Establishment Doppler receiving station is situated alongside the interferometer at Lasham, which gives opportunity of comparing the

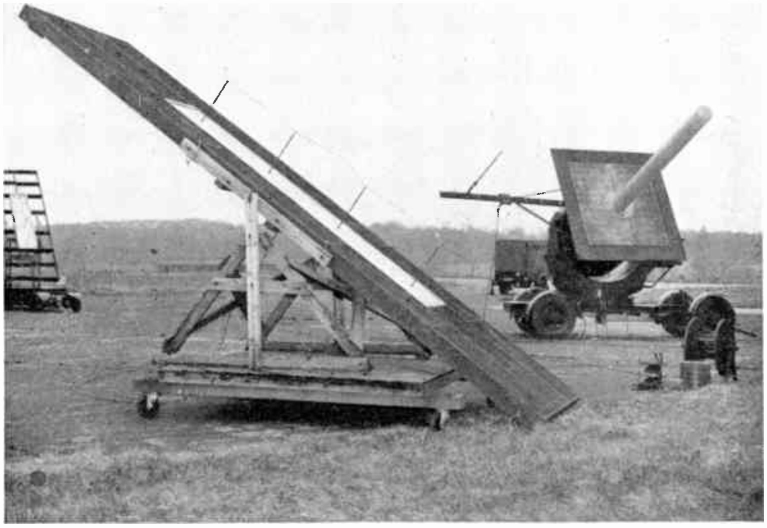


Fig. 5. Typical arrays of aerials for satellite observation. Yagi aerials, helical aerials, and arrays of elements of the type shown in Fig. 1 have been used in satellite observations.

two methods directly and, as in the interferometer case, observations can be made on various frequencies.

The signals are received on a variety of aerials, a group of which are shown in Fig. 5. These aerials are directional, and are controlled so as to point approximately to the satellite concerned by means of equipment shown in Fig. 6.

The main receiving equipment is shown in Fig. 7, in which one can see an accurate oscillator, receivers, timing equipment, and data recording equipment. The object of this equipment is to measure the received frequency with high precision at accurately known time. The data collected may be recorded, for subsequent processing, transmitted live to the main Establishment at Farnborough, or converted into suitable digitized form for transmission over standard teleprinter circuits. In our co-operation with the U.S.A. a transatlantic teleprinter circuit is used for transmission of data.

5. Some Sources of Error

In any radio method, one is dependent on transmission through the atmosphere and ionosphere. The radio waves are refracted, and this will cause both Doppler and interferometer readings to be in error.

The refraction error decreases with frequency and by choice of a sufficiently high frequency the error can be minimized. Further, by observing two frequencies simultaneously, an error correction factor can be found which is sufficiently good for any likely navigation purposes.

A second source of error may lie in the imperfections of the receiving sites. It is well known that obstructions cause errors in direction finding systems, including interferometers. For this reason the interferometers are installed on the best site to be found

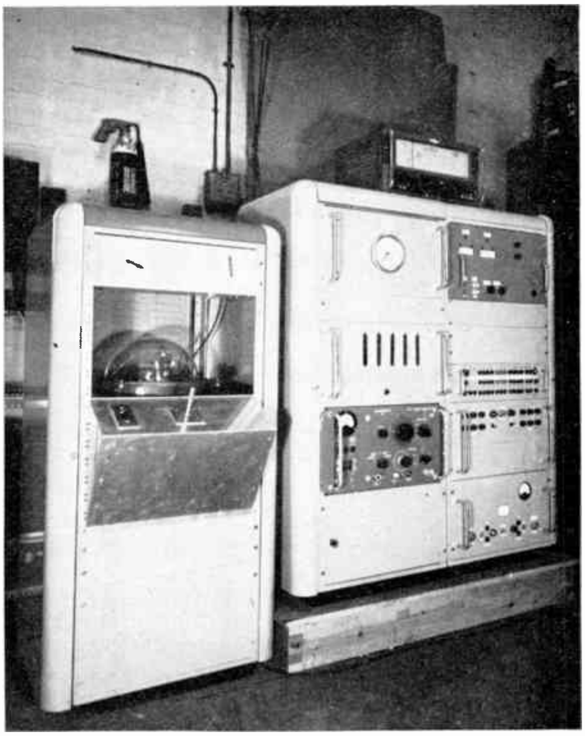


Fig. 6. Aerial control unit. The aerials shown in Fig. 5 may be pointed towards the satellite with the aid of this control unit.

A predicted track of the satellite is marked on the perspex dome, and a pointer is steered by hand along this track. The aerial mountings are servo controlled to follow the pointer. This is sufficiently accurate to track the aerials used.

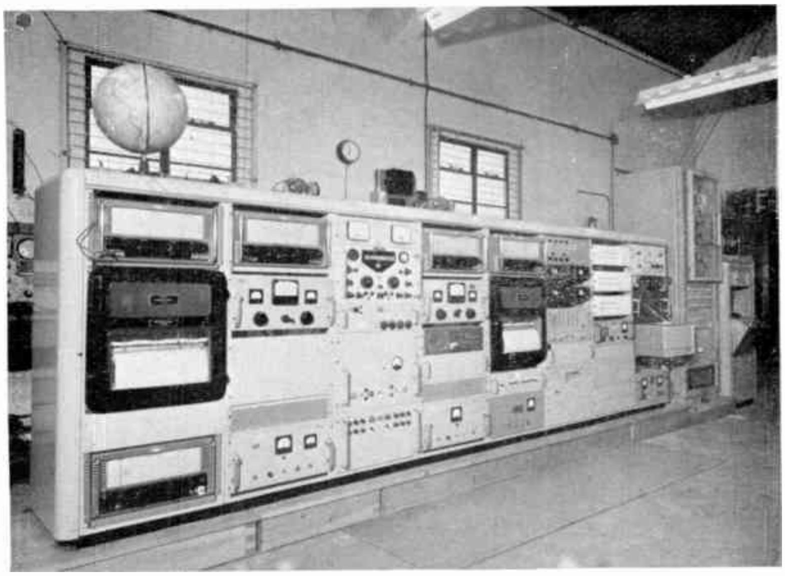


Fig. 7. The signals from the aerials in Fig. 5 are fed to the receivers at the top of the left hand racks, together with standard frequency signals generated in other racks. The heat frequency is recorded against time on the pen recorder together with field strength measurements. The Doppler signals are further processed and may be recorded for detailed analysis on the tape recorder on the extreme right.

at Lasham. While to a certain extent site errors may be calibrated out on a fixed orbit-determining station, the requirement for a good site, and the large area required, makes interferometers unattractive for a transportable system. However, we must not lose sight of the fact that if a signal is received simultaneously by two paths on a Doppler receiving station the Doppler curve can be distorted, and hence site errors can occur with the Doppler method.

6. Practical Results

In principle the methods of determining orbits, and using the orbit to determine positions are reciprocal. This was demonstrated at the Open Day at the Royal Aircraft Establishment in 1961, when signals received from the *Transit 2A* satellite at the station at Lasham were fed into the Establishment's Pegasus computer. Whereas the computer is normally used for the determination of orbits, on this occasion the published orbital data for *Transit* were used and the apparent position of Lasham was calculated. The position obtained was a mile or two in error, but one must realize that the published orbital data used are not quite up to date. This demonstration was, however, an indication that in an operational system with continuously updated, or more accurately predicted, orbital data, fixes of high precision should be possible. From the practical point of view there may be applications where a reliable fix to within a few miles is more important than the precision of the fix. In this case both the radio and computational aspects can no doubt be simplified.

7. Conclusions

Results of radio observations on satellites so far published not only give valuable scientific information on properties of the ionosphere but demonstrate that, given satellites in suitable orbits, navigational fixes can be obtained, the accuracy of which is limited principally by available orbital data. For many practical navigation purposes it should be possible to obtain approximate position fixes reasonably simply independent of weather or ionospheric conditions.

8. Acknowledgment

The author wishes to acknowledge the co-operation of the Applied Physics Laboratory, Johns Hopkins University, in supplying the illustration of the *Transit* system and the associated film which was shown during the Convention.

9. References

1. W. A. Scott, "Determination of position on the earth from a single observation of an artificial satellite", *Journal of the Institute of Navigation*, **14**, No. 1, pp. 87–93, January 1961.
2. Staff of the Royal Aircraft Establishment, Farnborough, "Observations on the orbit of the first Russian earth satellite", *Nature*, **180**, pp. 937–41, 9th November, 1957.
3. W. H. Guier and G. C. Weiffenbach, "A satellite doppler navigation system", *Proc. Inst. Radio Engrs*, **48**, pp. 507–16, April 1960.
4. J. D. Nicolaides, "Project TRANSIT," *Aerospace Engineering*, February 1961.

Manuscript first received by the Institution on 1st July, 1961, and in final form on 9th March, 1962. (Contribution No. 47).

Space Research Experiments in the World's First International Satellite

The world's first international satellite, containing space experiments devised and produced by British scientists, was successfully launched by the U.S. National Aeronautics and Space Administration (NASA), from Cape Canaveral, Florida, on 26th April 1962. The satellite, known previously as *Scout* 1, UK–1 or S–51, has been named *Ariel*. It contains equipment designed by scientists from University College London (in association with the University of Leicester), University of Birmingham, and Imperial College, London. The project is under the general direction of Professor Sir Harrie Massey, F.R.S. and is part of a co-operative programme arranged between the U.S. and British Governments.

Basically a cylinder 23 in. diameter and about 23 in. long, the satellite measures 10 ft across in orbit with instrument and aerial booms and solar cell paddles extended. The structure and ancillary equipment (telemetry and power supplies) were designed and constructed by NASA at their Goddard Space Flight Center.

The orbit of the satellite is inclined at an angle of 53·87 deg. to the Equator and it completes a revolution of the earth every 100·85 minutes; the apogee is 760 miles (1222 km) and perigee 236 miles (380 km).

Ariel is being tracked by the world-wide chain of NASA Minitrack stations, one of which is located in Berkshire and operated by the D.S.I.R.'s Radio Research Station at Datchet.[1] Other telemetry stations with command facilities have been set up at Singapore and Port Stanley, Falkland Islands.

The data from the experiments in the satellite are transmitted continuously (using the 137–137 Mc/s band) and at the same time some 1¾ hours of observations are stored in the satellite on a miniature tape recorder. This recorder can be commanded, by a radio signal from the ground, to play back the data at high speed to ground receiving stations. The data from one complete orbit can thus be received in a few minutes at one station.

For two or three consecutive orbits in each 24-hour period, the D.S.I.R. Radio Research Sub-station in the Falkland Islands is the only site suitably located for receiving these signals. Similarly, the D.S.I.R. Radio Research Station at Singapore can record transmissions not obtainable by any other station. It is planned that radio signals will be received for up to a year and that the satellite transmitter will be then permanently switched off.

The experiments, designed by the three University groups, have been selected for the satellite payload in consultation with NASA by the British National Committee on Space Research, set up by the Royal Society. Several of these experiments were described in papers read at the Brit.I.R.E. Convention in Oxford last year.

Dr. R. L. F. Boyd and his group at University College, London, have designed experiments to measure:

(i) Electron temperature and density (Langmuir probe)[2]

(ii) Ion mass composition and temperature (spherical probe)

(iii) Solar Lyman-α emission in the ultra-violet band

(iv) Solar x-ray emission in the 3–12Å band (in collaboration with the University of Leicester)[3]

(v) Solar aspect.

Professor J. Sayers and his group at the University of Birmingham have designed an experiment to measure electron density (using a radio frequency plasma probe). This experiment, whilst determining the same quantity as in Dr. Boyd's first experiment, does so by entirely novel methods and equipment. The data from both experiments will provide a cross-check on the variation of electron density at heights between 200 and 600 miles.

Professor H. Elliot and his group at Imperial College, London, have designed an experiment to measure the energy spectrum of the heavy primary cosmic rays.[4] This spectrum provides information about the earth's magnetic field and about the electro-magnetic conditions in interplanetary space.

Magnetic tapes bearing the data from the receiving stations are sent to the Goddard Space Flight Center, where the preliminary operations are carried out to process the data on to new tapes. These tapes will be further analysed on computers in this country, by the U.K. Atomic Energy Authority in co-operation with the Royal Aircraft Establishment at Farnborough. The data will then be in a form suitable for final analysis by the three experimental groups.

The second of the three satellites in the international programme, S–52, is scheduled to be launched by a *Scout* rocket from NASA's Wallops Station, Va., in 1963. It will be similar in construction to the S–51, and will be instrumented to conduct three main experiments:

(1) The measurement of galactic radio noise in the frequency range of 0·75 Mc/s and 3·0 Mc/s and exploration on the upper ionosphere.

(2) Measurement of the vertical distribution of ozone in the atmosphere.

(3) Measurement of micrometeorite flux with measurement of the quantity and size of the particles down to several microns in diameter.

References

1. "New satellite tracking station in Great Britain", *J. Brit.I.R.E.*, **21**, No. 2, pp. 150–2, February 1961.
   B. G. Pressey, "Radio tracking of artificial earth satellites", *J. Brit.I.R.E.*, **22**, No. 2, pp. 97–107, August 1961.
2. R. L. F. Boyd, "The use of probing electrodes in the study of ionosphere", *J. Brit.I.R.E.*, **22**, No. 5, pp. 405–8, November 1961.
3. K. A. Pounds, "Measurements of solar x-radiation", *J. Brit.I.R.E.*, **22**, No. 2, pp. 171–5, August 1961.
   J. Ackroyd, R. I. Evans and P. Walker, "X-ray spectrometer for Scout Satellite", *J. Brit.I.R.E.*, **23**, No. 1, pp. 55–60, January 1962.
4. H. Elliot, J. J. Quenby, D. W. Mayne and A. C. Durney, "Cosmic ray measurements in the U.K. Scout 1 satellite", *J. Brit.I.R.E.*, **22**, No. 3, pp. 251–6, September 1961.

U.D.C. 621.391.833 : 621.397.74 : 621.372

Picture Quality Assessment and Waveform Distortion Correction on Wired Television Systems

By

B. W. OSBORNE, M.Sc.

(*Associate Member*)†

Summary: Factors affecting the application of the *K*-rating as an assessment of picture quality on vestigial sideband television systems are mentioned, and the use of echoes for the correction of waveform distortion is briefly discussed. Description of the design and use of a compact, low-cost video frequency transversal equalizer is followed by consideration of an equalizer capable of correcting video waveform distortion on a modulated carrier. Using the latter, it was found possible to eliminate a long delay echo introduced at band I by the introduction of an echo at a lower "system" frequency.

1. The Practical Application of the *K*-rating to Vestigial Sideband Television Systems

It is appropriate first to begin with some comments on the measurement of television picture quality.

It is assumed that the reader is familiar with the work of Lewis[1,2] and Macdiarmid[3,4,5] on the measurement of subjective picture quality on linear systems in terms of the *K*-rating, using the pulse and bar test waveform; and with the advantages, in considering the performance of any television link, of thinking in terms of waveform distortion.[5] On a linear system, then, the picture quality can be directly assessed and quoted as a *K*-rating.

1.1. *Quadrature Distortion on Vestigial-sideband Television Systems*

Any television receiver, wired or otherwise, receiving a v.s.b. signal of high modulation depth, and employing an envelope detector, is not a linear device, for some quadrature distortion is inevitably present. Thus the proper application of the *K*-rating, for which the relevant subjective tests have been made only on linear systems, might be in doubt. Quadrature distortion has the greatest effect on the sine-squared pulse half-amplitude duration, and it is found that the *K*-rating is still a most useful measure of picture quality, provided that the half-amplitude duration limits are not enforced. In doing this, it should be realized that the subjective effects of relaxing the half-amplitude duration limits have not been determined, as the original *K*-rating subjective tests were made on linear systems.

It is thought that the pulse-widening effect of quadrature distortion on the 405-line positive modulation system is not directly associated with noticeable degradation of picture quality, since the change in half-amplitude duration is generally relatively small, and is not necessarily associated with any change in pulse height.

On a linear system, however, the area bounded by the sine-squared pulse waveform and the black level "baseline" is constant, so that any pulse widening is inevitably accompanied by a reduction in pulse height. A symmetrical change of this kind is of course typical of a system with a lack of high frequency response.

But the pulse widening associated with envelope detectors used on vestigial sideband systems with positive modulation and high modulation depths[6,7,8] is not necessarily associated with any loss of pulse height (or, in other words, it *is* possible to make television receivers on the 405-line system with adequate response at the higher video frequencies). The *K*-rating half-amplitude duration limits may thus be considered too stringent when applied to partial single-sideband systems.

Thus there is no full background of subjective measurements on which to base appropriate *K*-rating limits for commercial television receivers, though Macdiarmid[5] refers to the value of measurements on the 2*T* pulse, even in the presence of "substantial amounts of non-linearity".

1.2. *Proposal for Revision of K-rating Limits on Long-delay Echoes*

There is one aspect of the *K*-rating limits, the limits on echo amplitude over the delay range from 2 to 10 μs, where the *K*-rating does not appear to be a sufficiently stringent measurement of picture quality. Long delay echoes of very small amplitude are, under some circumstances, more visible than would be supposed from the point of view of their *K*-rating. Thus a 3% *K*-rating picture is normally a very good one, unless it is severely marred by the presence of a 3% long-delay echo.

It is proposed that the *K*-rating "pulse shape" limits at long time delays should be lowered, so that the 3% *K*-limit would correspond to something less than the 3% amplitude level at time delays beyond 8*T*. The 3% *K*-limit might well continue down to 1½%

† Rediffusion Research Limited, Kingston-on-Thames, Surrey.

affects both the delay/frequency and the amplitude/frequency responses. Further reference to a practical application of this will be made later.

Alternatively, we can introduce echoes in order to obtain a pre-determined amplitude/frequency or delay/frequency characteristic. A transversal filter may be used, e.g. to generate vestigial sideband signals with no delay distortion,[11] or, very simply, by the use of a single echo pair, to demodulate frequency-modulated television signals,[13] or, generally, to obtain a desired amplitude response with no delay distortion.

It is likely that the application of transversal filters to television systems, and in particular to colour systems, has not yet been fully developed.

3. A Simplified Video Frequency Transversal Equalizer

Though transversal equalizers using many echo pairs and capable of very high performance have been used for the correction of long-distance television links[2,15] these equipments were generally too bulky and costly for general use on wired television distribution systems. At the same time, there was a need for a device to correct waveform distortion over a certain limited time scale.

It was found that on television distribution systems, such as that described by Kinross,[17] any correction needed was confined to the vicinity of a single picture element. It was decided to engineer an equalizer, primarily intended for use at a receiving site, but which could be used on the television network wherever the signal is brought down to video frequency.

It will be realized that in practice it is seldom desired to correct the amplitude/frequency and the delay/frequency characteristics separately. Many unwanted distortions, e.g. one due to a slight impedance mismatch, themselves produce sinusoidal terms of delay and amplitude simultaneously, equivalent to those introduced by a single echo; and they can be corrected by the introduction of a single echo arranged in amplitude and time so as to give an equal and opposite effect. Thus it is not generally necessary to use symmetrically spaced echo pairs.

The maximum available echo amplitudes at various time delays should not be equal, as echoes of given amplitude at greater time delays have more effect on subjective picture quality than those closer in. In this respect it will be noted that the amplitude of the harmonic terms of delay, due to an echo, is proportional to the product of the echo amplitude and the echo delay.

A video frequency transversal equalizer has been designed which introduces "positive" or "negative" echoes, each of adjustable amplitude, at time intervals of $-T$, $+T$, $+2T$ and $+3T$ simultaneously. These echoes can then be added to the original signal in such a way as to cancel the unwanted distortion and thereby improve the picture quality.

The equalizer is contained with its power supply on a standard rack mounting panel of 5¼ in. height; a block diagram is shown in Fig. 3. The incoming video frequency signal, at 1 volt level and 75 ohms impedance, is fed through a buffer stage to the input of a 600 ohm tapped delay line. The output from the first tapping point, at time t_0, is fed through an amplifier to the output cathode follower.

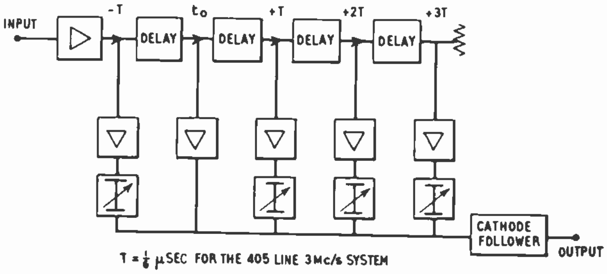


Fig. 3. Block diagram of a simple video frequency transversal equalizer.

The output signal is also made up of contributions obtained from amplifiers fed from tapping points on the delay line at times $-T$, $+T$, $+2T$ and $+3T$ respectively, relative to t_0, these contributions being controlled in both amplitude and polarity by the controls on the front panel.

If the echo amplitudes for the $-T$, $+T$, $+2T$ and $+3T$ echoes are A_1, A_2, A_3 and A_4 respectively (e.g. for 5% echo $A = 0{\cdot}05$) then for small echoes the resultant envelope delay for any frequency is given approximately by

$$\frac{\mathrm{d}\beta}{\mathrm{d}\omega} = t_0 - A_1 T\cos T\omega + A_2 T\cos T\omega + \\ + 2\,A_3\,T\cos 2\,T\omega + 3\,A_4\cos 3\,T\omega$$

where t_0 is constant; and the relative attenuation α for any particular frequency is given approximately by

$$\alpha\,(\mathrm{d}\beta) = 8{\cdot}69\,(A_1\cos T\omega + A_2\cos T\omega + \\ + A_3\cos 2\,T\omega + A_4\cos 3\,T\omega)$$

It is not generally necessary, or even useful, to evaluate these expressions for particular echo amplitudes, where the equalizer is set up for minimum waveform distortion. This further emphasizes the relative simplicity of the waveform distortion approach to television transmission systems.

The time interval T is of course dependent on the upper limit of the video frequency band in use on the particular television system. It is thus a very simple matter to change the time intervals, for different systems, without otherwise changing the equalizer design.

The equalizer has been found to be simple to operate, and, within its time scale, very flexible. It can, for instance, introduce amplitude/frequency correction (corresponding to a single cosine term, as mentioned in Section 2.3 above) by setting the levels of the $+2T$ and $+3T$ controls to zero, with the $-T$ and the $+T$ amplitude controls set to equal levels, the $-T$ and $+T$ echo polarities being both positive for a loss of high frequency response, or both negative for an increase in high frequency response. Generally, of course, the controls are set for minimum K-rating when viewing a sine-squared pulse and bar waveform, and the actual amplitude/frequency characteristic of the system need not be known.

The unit was made using six similar valve envelopes, each being a particular industrial type long-life double triode, working off an h.t. rail of 100 volts. Units have so far been made for use on the 525 line system in Canada, and on the British 405 line system.

4. The Correction of Video Waveform Distortion on a Modulated Carrier

Consider first the introduction of a single video frequency echo of amplitude A at a time delay T.

This produces a phase change of $A \sin T\omega$ radians, and thus a change of envelope delay of $AT \cos T\omega$. (This delay change is positive for a positive-going delayed echo, and negative for a positive-going preceding echo.)

If A and T are both variable, we can obviously obtain a wide range of sinusoidal variations of delay with frequency.

The corresponding change in attenuation α with frequency is given approximately by

$$\alpha = 8{\cdot}69\, A \cos T\omega \text{ decibels}$$
(for small echo amplitudes).

If we now apply a modulated carrier to a single-echo equalizer, we find that, as we vary the echo dẹlay T and view the demodulated video output, the echo reverses video polarity as T is moved through half the reciprocal of the carrier frequency. Thus, in order to obtain proper control of the equalizer, it is essential to provide steps of echo delay time that are sufficiently small. Steps not greater than about 30 degrees of carrier phase are convenient.

If the carrier frequency is comparable with the high modulation frequencies, as is often the case on wired television distribution systems[17] it will be noted that, as the echo delay produces different degrees of phase shift at the different frequencies, the setting of echo delay time will also, for example, affect the pulse-to-bar ratio on the sine-squared pulse and bar test waveform.

Though it is generally more convenient to use a transversal equalizer at video frequency in the manner described in Section 3 above, it has often been the practice, on wired television networks, to generate the system-frequency carrier by mixing down from the transmitted frequency. Thus where the received signal is not demodulated before reaching the subscriber's wired receiver, there is a need for a device to correct for video waveform distortion by operating on the modulated system-frequency carrier.

At the comparatively low system frequencies used, it is fortunately the case that the design of a single delay line section giving a sufficiently short step of time in terms of carrier phase also gives a sufficiently high cut-off frequency. Whereas for the video frequency equalizer of Section 3 it was possible to use lengths of suitable delay cable, for operation on modulated carriers it was found necessary to make up lumped-circuit delay line sections.

In the equalizer designed for this purpose each delay section is made to introduce 10 millimicroseconds, at a line impedance of 100 ohms. Sixty of these sections provide a maximum time delay of 600 millimicroseconds, half of this delay being mounted on a suitable rotary switch controlled by a large knob in the centre of the panel. The equalizer is also provided with a phase reversal switch (as without this the correct video delay might correspond to an out-of-phase carrier relationship, giving a black echo where a white one was required), and a suitable echo attenuator.

5. The Elimination of Long Delay Echoes

The equalizer of Section 4 has also been made, in a modified form, for dealing with a single long-delay echo. On some aerial sites, particularly in hilly country, it may not be possible to eliminate all unwanted reflections by choice of aerials, and it is then desirable to remove the offending echo by introducing one equal and opposite.

It was mentioned in Section 1 that it is thought that the "pulse shape" K-rating limits are not sufficiently stringent at long delay times; this aspect of the practical interpretation and use of the K-rating first became apparent in the correction of very weak long-delay echoes in the Rhondda in South Wales, where the hills and slag-heaps make it particularly difficult to eliminate all unwanted echoes at the aerial site. Investigation showed that long-delay echoes as small as 2% or 3% were visible on test cards, or on suitable programme material (white captions on a black background, etc.). These very small echoes constituted a nuisance which would not be tolerated by subscribers, even when viewed on wired receivers of not better than 5% K-rating.

An external delay line unit, providing an additional time delay of 3 μs in 0·5 μs steps, was made so as to extend the time scale of the equalizer described in Section 4, which could then be adjusted in steps of 10 mμs up to 3·6 μs. (Delays as long as 10 μs have been used in the laboratory.)

The use of this long-delay echo equipment has made it possible to remove an unwanted echo, caused by some unavoidable reflection of the Band 1 transmitted signal, by the introduction of an echo after frequency conversion to the 4·95 Mc/s system frequency. It would have been completely impracticable to construct such long delay lines to work at Band 1, whilst it would have been inconvenient and costly to come down to video and re-modulate merely for the purpose of removing the echo. However, it must be remembered that it is only possible to remove an echo in this way if the echo is fixed in delay and polarity.

6. Some Further Remarks on the Application of Waveform Testing

The use of waveform testing methods on wired television networks is now well established,[10] and it is thereby possible to assess the overall performance of a network, right up to and including the subscriber's wired receiver, in terms of subjective picture quality. Further, the use of the same methods enables the broadcasting authorities to control picture quality from the studio to the transmitter.

As line-repetitive test waveforms from the studio are not generally broadcast, there is as yet no complete information on subjective picture quality in terms of the distortion of the test waveforms over the complete link from studio to viewer. A pulse-and-bar signal inserted in the frame blanking interval is not adequate in practice for accurate measurement of waveform distortion, mainly because of the poor trace brightness obtained on most c.r.t. displays when viewing a spot moving at about 10 cm/μs across the screen at a repetition frequency of 50 c/s.† Thus, unfortunately, it is generally necessary to adjust the equalizers of Sections 3 and 4 above when viewing a transmitted test card or other picture, instead of on the much more sensitive and reliable pulse-and-bar waveform display.

If it is assumed that the broadcast signal is effectively undistorted (and this cannot be supposed to be generally true at differing receiving sites) there must still be some advantage (not restricted to wired television distribution systems) in having a transmitted line-repetitive test waveform (perhaps for half an hour, before programme hours) in order to monitor subjective picture quality, in terms of the pulse-and-bar waveform, at the receiving site or anywhere on the distribution network. The licensing authorities would also find it easier to assess the performance of the various wired television systems in the matter of picture quality.

† Also, the frame inserted signal may not always be acceptable to viewers using older receivers.

7. Acknowledgments

Thanks are due to Dr. N. W. Lewis and Mr. I. F. Macdiarmid, of the Post Office Research Station, for several most useful discussions; to those at Rediffusion Research who have taken part in the work described here; to the Directors of Rediffusion Research Limited, by whose permission this Paper is published, and in particular to the Managing Director, Mr. R. I. Kinross.

8. References

1. N. W. Lewis, "Waveform computations by the time-series method", *Proc. Instn Elect. Engrs*, **99**, Part III, pp. 294–306, 1952.
2. N. W. Lewis, "Waveform responses of television links", *Proc. Instn Elect. Engrs*, **101**, Part III, pp. 258–70, 1954.
3. I. F. Macdiarmid, "A testing pulse for television links", *Proc. Instn Elect. Engrs*, **99**, Part III A, pp. 436–44, 1952.
4. I. F. Macdiarmid, "Waveform distortion in television links", *Post Office Elect. Engrs J.*, **52**, pp. 108–14 and 188–95, 1959.
5. I. F. Macdiarmid, "Waveform distortion in television links", *J. Brit. I.R.E.*, **20**, pp. 201–16, 1960.
6. H. Nyquist and K. W. Pfleger, "Effect of the quadrature component in single sideband transmission", *Bell Syst. Tech. J.*, **19**, pp. 63–73, 1940.
7. R. D. A. Maurice, "Comparison of four television standards", *Electronic Radio Engr*, **34**, pp. 416–21, 1957.
8. R. D. A. Maurice, Correspondence on reference 7, *Electronic Radio Engr*, **36**, pp. 352–3, 1959.
9. P. Mertz, "Influence of echoes on television transmission" *J. Soc. Mot. Pict. Telev. Engrs*, **60**, pp. 572–96, 1953.
10. B. W. Osborne, "Picture quality control equipment for wired television networks", *Proc. Soc. Relay Engrs*, **5**, pp. 89–119, 1961.
11. A. D. Blumlein, H. E. Kallman and W. S. Percival, British Patent No. 517,516, 1938.
12. H. A. Wheeler, "The interpretation of amplitude and phase distortion in terms of paired echoes", *Proc. Inst. Radio Engrs*, **27**, pp. 359–85, 1939.
13. H. E. Kallman, "Transversal filters", *Proc. Inst. Radio Engrs*, **28**, pp. 302–10, 1940.
14. J. M. Linke, "A variable time-equalizer for video-frequency waveform correction", *Proc. Instn Elect. Engrs*, **99**, Part III A, pp. 427–35, 1952.
15. R. V. Sperry and D. Surenian, "A transversal equalizer for television circuits", *Bell Syst. Tech. J.*, **39**, pp. 405–22, 1960.
16. A. D. Fowler and J. D. Igleheart, "Effects of frequency cut-off characteristics on spiking and ringing of tv signals", *Trans. Inst. Radio Engrs (Communication Systems)*, **CS-7**, pp. 173–9, 1959.
17. R. I. Kinross, "Distribution of television by wire", *Proc. Soc. Relay Engrs*, **4**, pp. 13-41, 1957.

Manuscript first received by the Institution on 11th May 1961 and in final form on 16th August 1961. (*Paper No. 729*).

U.D.C. 621.385.44.029.63

A Rugged 3 kMc/s, 40-Watt Transmitting Tetrode

By

J. J. HAMILTON, M.Sc.(Eng.)
(*Associate Member*)†

Summary: A planar tetrode developed primarily for the amplification of r.f. power in the region of 3 kMc/s, is described. The valve comprises a rugged, thermally stable electrode structure capable of providing 40 watts of c.w. output power at 3 kMc/s with a gain of 7·5 dB. Some unusual design features of the valve and relevant performance data are presented.

1. Introduction

The rapid growth of microwave communication and the stimulus of advances registered in the areas of velocity-modulated, crossed-field and travelling-wave devices, have for some time presented a challenge to engineers engaged in the design of space-charge valves.

Novel and successively bolder design approaches, motivated by these developments, have, in turn, given a new lease of life to the state of the art of space-charge valves intended for microwave operation. The valve described in the present paper constitutes one such development,‡ undertaken with the object of introducing medium power space-charge amplification to the region of 3 kMc/s. The project also had as secondary goal the attainment of high standards of environmental performance in the valve.

At 3 kMc/s the design of space-charge valves is largely governed by factors which are of little consequence at lower frequencies. In order to satisfy restrictions imposed by operation at microwave frequencies, the design of space-charge valves must in some ways depart from established valve design criteria pertaining to lower r.f. work. Hence, the parameters of the resulting valve are orientated primarily toward the accomplishment of a specific function in the field of microwave applications.

Such is the case of the General Electric valve type Z-5267, shown in Fig. 1, whose principal design features are described in the following sections.

2. Electrical Design Considerations

The advantages inherent in a power tetrode, namely:

(1) r.f. input-output isolation,

(2) higher gain at the desired power level, and

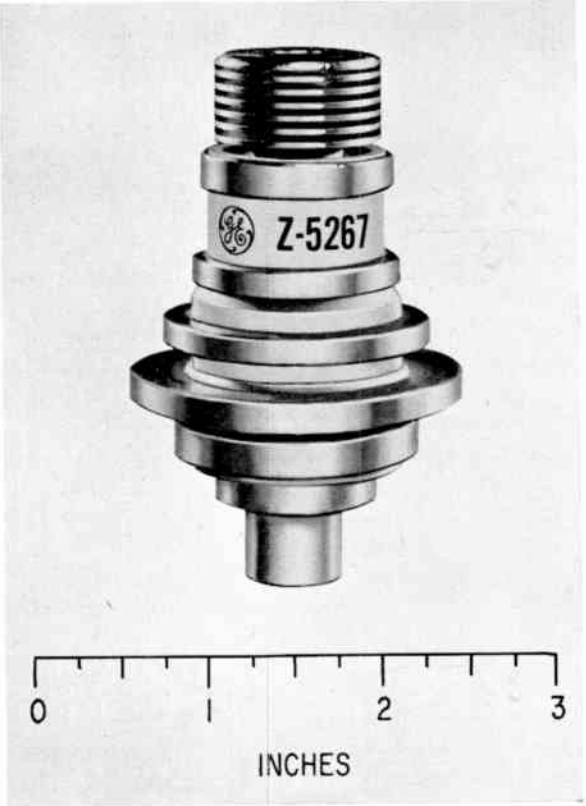


Fig. 1. Valve type Z-5267—3 kMc/s transmitting tetrode.

(3) more favourable transit time conditions, hence better efficiency,

determined its selection over a power triode for the intended frequency and power level of operation.

In the region of normal operation, the anode current of a tetrode is primarily determined by the control and screen grid voltages. As far as the electron flow is concerned, therefore, the cathode, control and screen grid region of the tetrode constitutes an input triode section, serving to modulate and accelerate the electron beam, which may be optimized independently of the rest of the valve.

† General Electric Company, Power Tube Department, Schenectady, N.Y.

‡ Sponsored by U.S.A.S.R.D.L. under Contract No. DA 36-039-sc-75069.

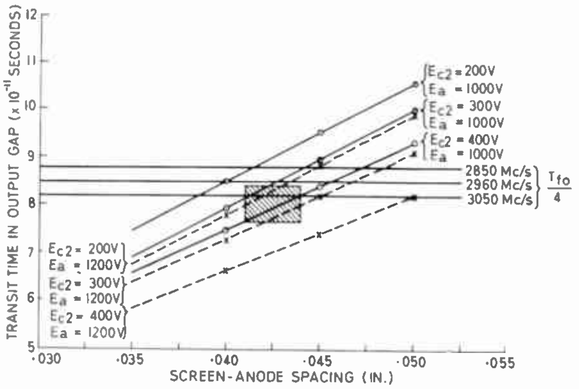


Fig. 2. Output gap transit time as a function of spacing and applied voltages.

The screen-anode region, on the other hand, serves the purpose of a gap which demodulates the beam and thereby transfers r.f. energy from the beam to the output load. This region may therefore be optimized independently also. Ultimately, mutual compatibility of the two sections must be attained, and their strict compliance with established criteria of sound valve design secured, through successive approximation.

Theoretical analyses of triodes have been reported by several workers in the field.[1,2,3] The input triode section of the Z-5267 was designed on the basis of similar treatment, with particular emphasis placed on inter-electrode transit time, high transconductance and adequate power handling capability. The control grid-cathode spacing was set for a transit angle of less than π radians while the control-screen grid spacing was kept purposely small to reduce the effect of "smearing" on the modulated electron beam.[4] The obvious advantage of a low input gap capacitance favoured the choice of the smallest cathode button feasible in terms of operation at a safe current density level. The grid wire diameter and pitch were selected after due consideration of transconductance, and the physical limitations of fine diameter wires.

The design of the output gap was controlled by several factors. One of these, the anode diameter, was fixed by the size of the proposed cathode. As the latter was already designed for least active surface area however, the requirement for low output gap capacitance was not seriously handicapped by this restriction. In actual fact, the diameter of the anode was made slightly larger than that of the cathode to safeguard against damping of the output gap by stray electrons. The spacing of the gap was chosen to comply with a maximum anode voltage requirement of 1000 volts and a proportionate level of screen potential. It was designed to provide an electron transit time of a quarter of a cycle or less for efficient interaction between the beam and the output gap. It was also calculated to preclude the formation of a potential minimum which would have the effect of lengthening the electron transit time in the screen-anode region, thereby imposing a drastic reduction in the interaction efficiency of the output gap.

A serious conflict of interests arose in dimensioning the high-dielectric ceramic cylinder surrounding the gap. Here, substantial wall thickness was prerequisite to adequate mechanical strength. Compensating for this, the inner diameter of the ceramic had to be made large enough to offset any undue capacitive loading of the gap. At the same time, however, its outer diameter had to be of restricted size to permit the design of a practical external 3 kMc/s resonator. In short, it was necessary to arrive at an output gap configuration which would meet stringent mechanical, as well as both internal and external r.f. requirements.

The above factors were examined collectively and some direct results are shown plotted in Fig. 2. The quarter period corresponding to the specific resonant frequencies quoted is represented by $T_{fo}/4$. The transit time of the slowest electrons is considered, and the shaded area represents the region wherein all conditions pertaining to the design of the output gap are, to a greater or lesser extent, satisfied.

The screen is wound at half the number of turns per inch used on the control grid since electrostatic plots showed this to have little effect on the field pattern of the control grid-cathode region. The ensuing reduction in screen current is of benefit to both the screen power supply requirements and the operating efficiency of the valve. Furthermore, the control grid is wound with larger diameter wire than the screen

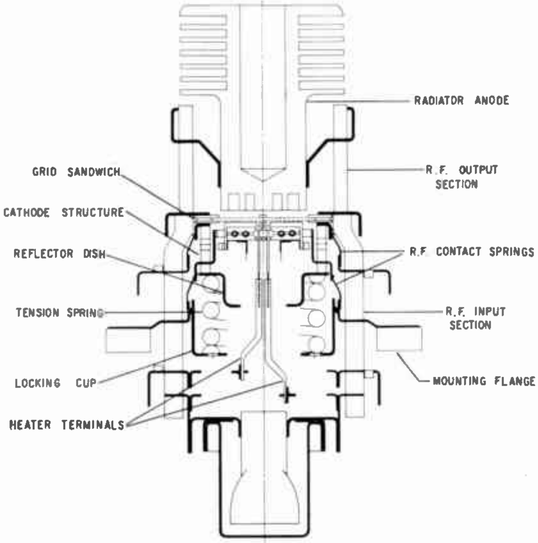


Fig. 3. Z-5267 layout drawing.

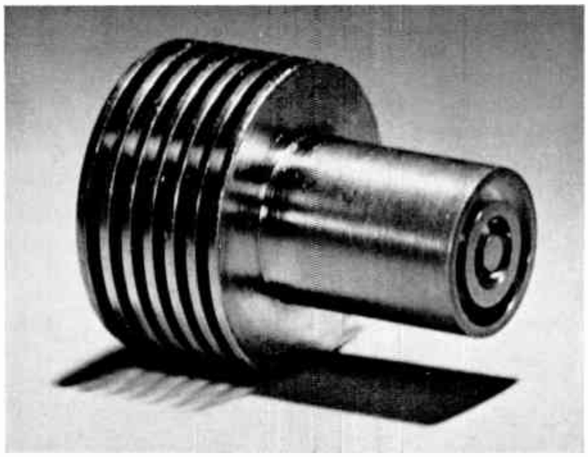


Fig. 4. Grooved radiator anode.

to promote improved shadowing of the latter. It was found during the course of development of the Z-5267 that, with equal wire diameter grids, the screen would, at best, remain operative only up to an r.f. power output level of 35 watts at 3 kMc/s. The crossbars carried by the grids (Fig. 6), which necessitate slotting of the cathode, loss of emissive surface area, and the introduction of redundant capacitance, were required to provide the valve with a rugged, thermally stable electrode structure.

Whereas no individual parameter is fully optimized in the Z-5267 tetrode, a balance has been obtained wherein all are essentially satisfied, and the desired performance attained at 3 kMc/s.

3. Structural Features

The objective requirement for a valve capable of withstanding high impact shock of 450 *g* and operation at 300° C body temperature necessitated the design of a metal-ceramic structure, as shown in the layout drawing of Fig. 3.

A butt-seal envelope, consisting of largely self-jigging high percentage alumina ceramics and copper-clad stainless steel parts, was chosen both because of its inherent ruggedness and the relative ease attending the manufacture of stacked assemblies. Concentric cathode, control grid, screen grid and anode terminal ring connectors are stacked in order of decreasing circumference, thereby providing a valve with single-ended resonator terminals for "plug-in" convenience.

An integral copper radiator anode is incorporated in the valve for efficient heat transfer. The central portion of the radiator is hollowed out in the interests of weight reduction, since its contribution to the net heat flow is relatively small. The anode face itself (Fig. 4) is grooved as a means of suppressing deleterious secondary electron emission in the area of the output gap. The grooves are designed for the minimum depth considered effective in the suppression of secondaries in the screen-anode region of the valve. This serves the purpose of reducing the risk of long path discharge under severe overload conditions. The internal r.f. input contact rings, shown anchored to the cathode and control grid flanges in Fig. 5, are designed to provide the quality of r.f. contact necessary at the power level and frequency of intended operation of the valve.

Also in keeping with the objective requirements, the grid sandwich and cathode structure contain novel features which enhance their ruggedness and thermal stability. The grid sandwich shown in Fig. 6 consists of a control grid, a screen grid and an isolating, high dielectric, alumina ceramic disc constituting an effective internal grid by-passing element. The grids are wound on integral tungsten washers comprising diagonally slotted crossbars. These crossbars are designed to provide added support against wire sag and to raise the natural resonant frequency of the individual wires while allowing for free self-expansion during hot operation of the valve. The control grid is gold-plated to safeguard against primary electron emission. The screen retains a thin film of copper acquired during the wire brazing operation. This is considered adequate for purposes of r.f. conductivity. Following precise visual alignment of the grid wires, the separate parts are brazed into a rugged grid sandwich assembly.

The cathode assembly is made up of a cathode structure (Fig. 7) and a heater assembly (Fig. 8) installed in the slotted cathode button of the former. The cathode structure represents a new step in ruggedness. A pair of formed straps welded alternately to the cathode button and its supporting cylinder, 90

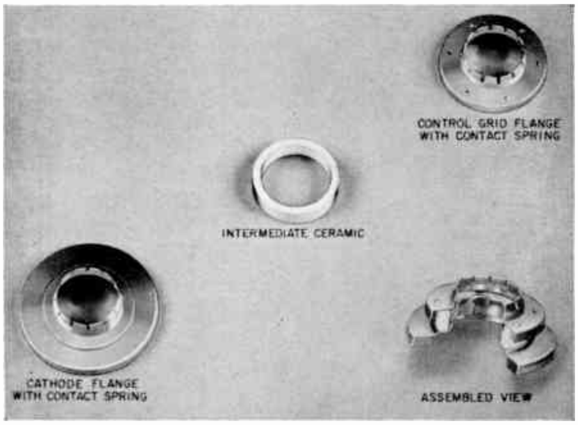


Fig. 5. Input r.f. contact springs.

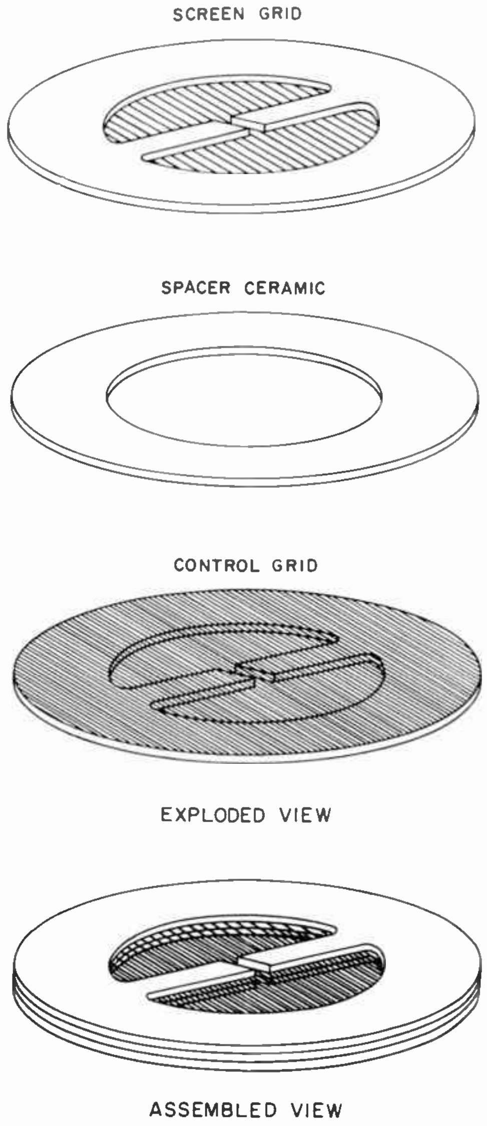


Fig. 6. Grid sandwich.

degrees out of phase with respect to each other, provide positive four-point support. This structure shares the high-heat-insulating properties of cathodes using cylindrical foil support. Warm-up displacement of the cathode face however is very limited, for as the straps expand, they bow laterally without imparting axial movement to the cathode. An efficient, ceramic-sandwiched pancake heater, whose separate parts are displayed in Fig. 8, provides another rugged feature of the Z-5267 cathode. The grooved ceramics locate the heater permanently within the cathode button. A titanium cylinder welded to the heater package provides effective gettering action in the valve. The unslotted portion of the cathode button is coated with a conventional triple carbonate mix.

After referencing the grid sandwich with respect to the cathode structure, the critical grid-cathode spacing is checked. The resulting sub-assembly is then inserted into the valve envelope, covered with a heat reflector and locked in position by the use of a compression spring and locking cup mechanism. This is done to preserve the small interelectrode spacings of the valve through subsequent environmental punishment. Heater lead support rings are welded to

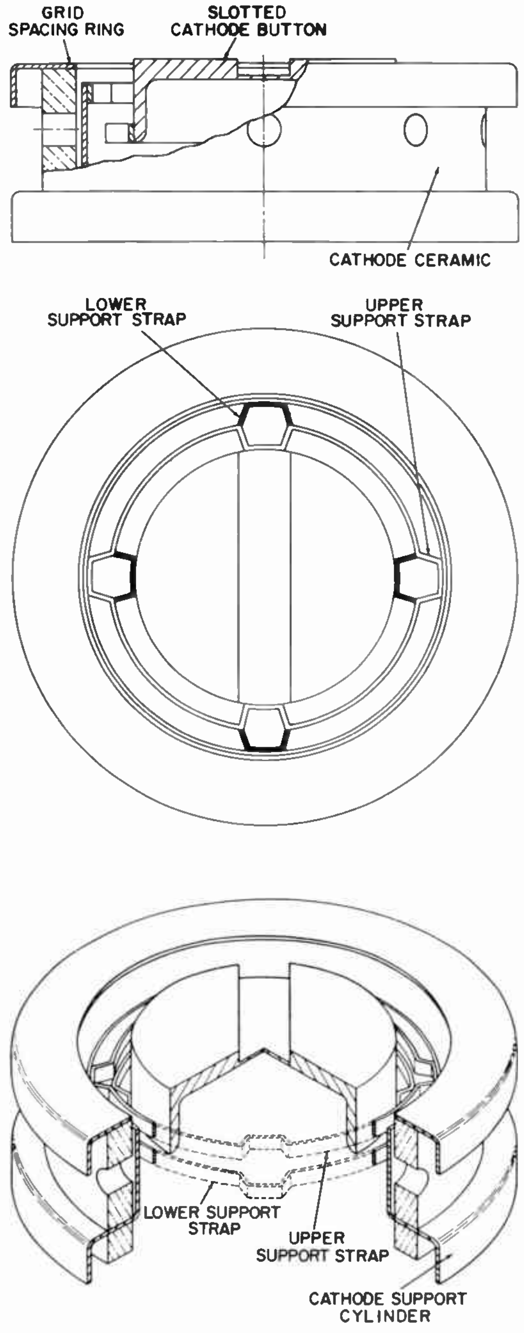


Fig. 7. Cathode structure.

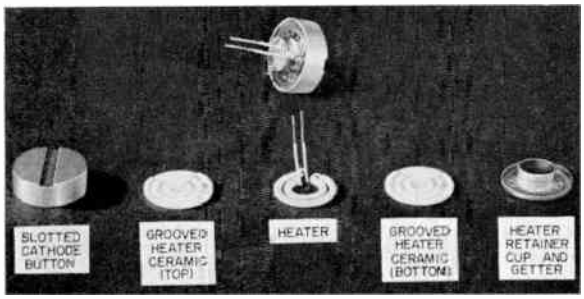

SLOTTED CATHODE BUTTON
GROOVED HEATER CERAMIC (TOP)
HEATER
GROOVED HEATER CERAMIC (BOTTOM)
HEATER RETAINER CUP AND GETTER

Fig. 8. Heater components and assembly.

the envelope heater flanges and the heater legs are in turn welded to their respective terminal crosspieces. The tube is silver plated following heliarc-welding in place of the exhaust cup and tubulation.

During exhaust the valve is baked out at 550° C and the oxide cathode is subjected to fast breakdown. After seal-off the tubular end cap is resistance-welded to its anchoring cup to obviate thermal limitations introduced by the use of solder.

The valve is 3 inches long and weighs approximately 4 ounces. Its largest diameter, represented by the cathode flange, stands at $1\frac{3}{4}$ inches.

4. Performance

4.1. *Electrical*

The principal design factors and operating parameters of the developmental valve are listed in Table 1. Its static characteristic curves are presented in Fig. 9.

Evaluation of the performance of the Z-5267 as a power amplifier at 3 kMc/s was realized through the design of external cavities for the input and output sections of the valve. The complete 3 kMc/s amplifier cavity system developed for this purpose is shown in Fig. 10.

A box-type resonator tuning over a range of several hundred megacycles by means of adjustable tuning plungers was designed for the input portion. A fixed probe was used to provide the necessary input coupling.

The required anode load impedance was supplied by the design of a suitable quarter wavelength radial cavity. D.c. isolation between the screen, the peripheral wall of the cavity and the anode of the valve was provided by the interposition of a set of quarter wavelength radial lines. A metallic tuning slug was included to furnish a limited mechanical tuning range of 18 Mc/s in the output cavity. Variable anode load coupling was made possible through the provision of a coupling loop of adjustable depth of insertion.

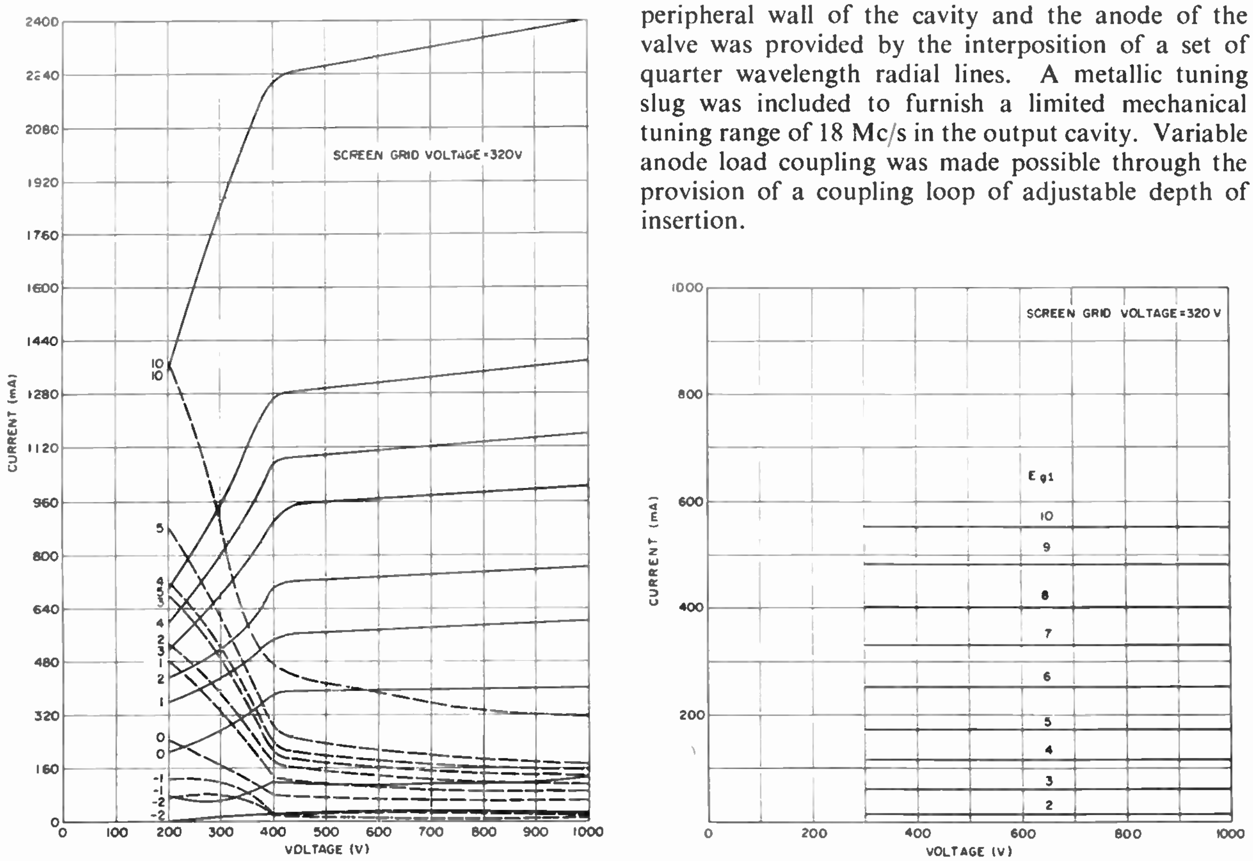


Fig. 9. Static characteristics of Z-5267.

Table 1

Details of the Z-5267 Tetrode

Typical Valve Constants	
Cathode diameter	0·43 in
Useful cathode area	0·115 in^2
Control grid—cathode spacing	0·0028 in
Control grid—screen grid spacing	0·015 in
Screen grid—anode spacing	0·0425 in
Control grid wire diameter	0·0012 in
Distance between control grid wires	0·004 in
Screen grid wire diameter	0·0008 in
Distance between screen grid wires	0·008 in
(Measurements are to centres of grid wires.)	
Direct interelectrode capacitances (approximate without shielding):	
Screen grid-to-anode	4 pF
Control grid-to-cathode	18 pF
Control grid-to-screen grid	30 pF
Amplification factor, g_1-g_2	60
Transconductance	60,000 μmhos
Typical Operating Constants	
Anode voltage	1000 V
Anode current	160 mA
Cathode current density	260 mA/cm^2
Screen grid voltage	300 V
Control grid voltage (approximate)	−1·2 V
Heater voltage	6·3 V
Heater current	1·9 A

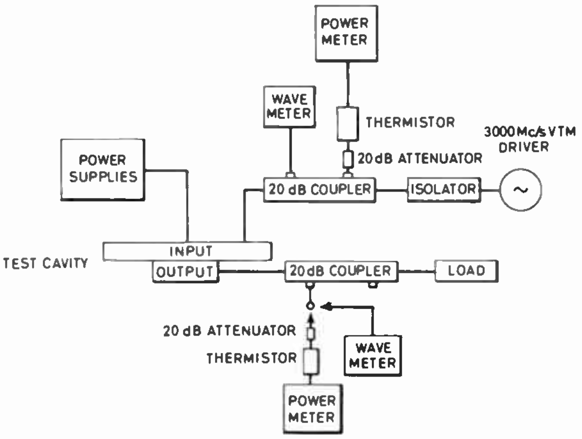


Fig. 11. Equipment used in 3 kMc/s amplifier test.

The amplifier cavity system incorporates external grid by-pass capacitance. This, in addition to the capacitance included between the screen and control grids as a feature of the Z-5267 tetrode, is intended to enhance efficient grounded-grid operation.

A block diagram of the test circuitry employed to obtain the 3 kMc/s test results is given in Fig. 11. Results obtained from tests performed using the above equipment indicated that the valve was capable of stable and efficient power amplification at this frequency. Representative curves of power output as a function of input and of variation in valve gain and efficiency with power output are included in Figs. 12 and 13.

4.2. *Environmental*

The environmental behaviour of the valve surpassed expectations as far as shock and high temperature goals were concerned. At the same time the basic vibration requirements were adequately met.

The tube was shock tested at 450 *g* in accordance with MIL—Specification instructions, without visible exterior damage or discernible change in basic valve characteristics. In this test, the valve was supported by its cathode mounting flange which was clamped peripherally in a special test fixture.

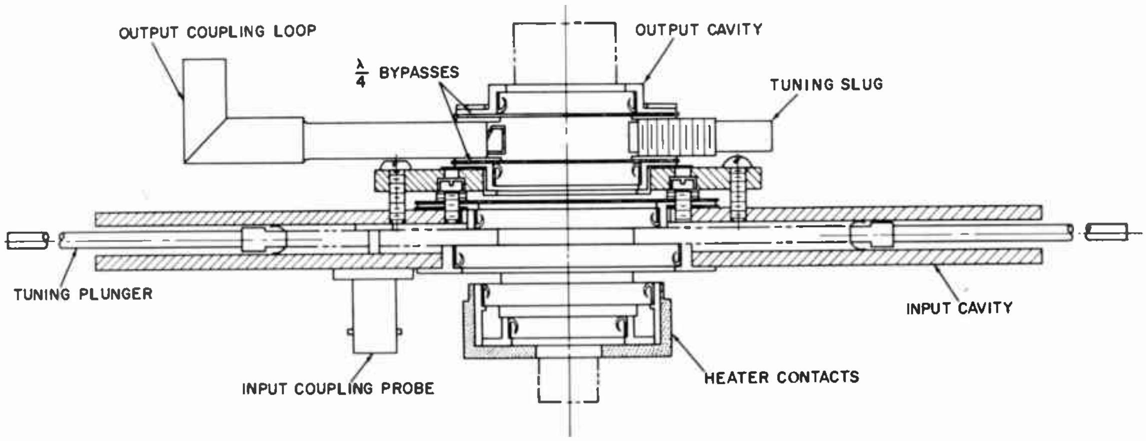


Fig. 10. 3 kMc/s power amplifier cavity system.

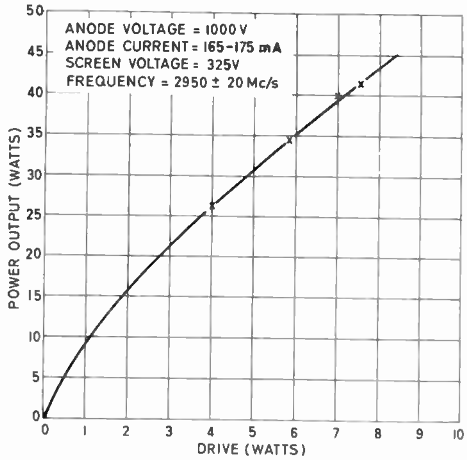


Fig. 12. Power output characteristic of Z-5267 tetrode.

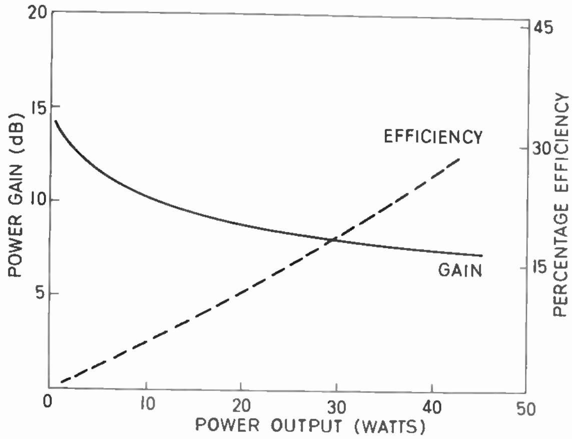


Fig. 13. Variation in gain and efficiency with power output.

Following the development of a non-resonant vibration socket for the frequency range 50 to 1600 c/s, the valve was subjected to vibration testing in this range to a maximum acceleration of 15 *g*. The noise output of the valve was measured across a non-inductive anode load resistor of 1500 ohms, at an anode current flow of 100 mA. An average noise output of 500 millivolts r.m.s. was observed, with sporadic fluctuations not exceeding 3·5 volts r.m.s. occurring in the region of 1000 to 1600 c/s. No change of a permanent nature was noted in the valves subjected to vibration testing.

The valve also functioned normally, without decline in cathode emission level, at body temperatures of up to 300° C.

4.3. *Life*

R.f. life tests conducted on Z-5267 valves for 1000 hours in the region of 3 kMc/s and 40 W useful output showed no sign of deterioration in the electrical characteristics of the valve up to that point.

Similarly, no loss in performance was observed when valves were subjected to operation at 300° C body temperature for periods ranging between 100 and 1000 hours.

5. Conclusion

Valve type Z-5267, a tetrode of planar construction, has been developed especially for the purpose of providing efficient medium power space-charge amplification in the region of 3 kMc/s. To this end, it combines small inter-electrode spacings with effective r.f. isolation between its input and output gaps and a thermally stable electrode structure. Operating at electrode voltages no greater than 1000 V, and utilizing a conventional oxide-coated cathode run at the higher current density levels encountered in transmitting lighthouse triode practice, the valve is capable of delivering 40 watts of output power at a gain of 7·5 dB. At this level of operation, the overall tube and circuit efficiency is in the region of 25% and the half-power bandwidth 25 Mc/s.

Lightweight and extremely rugged, the Z-5267 may be operated at body temperatures of up to 300° C, and has a life expectancy of several thousand hours. The tube is therefore eminently suited to operation as c.w. power amplifier in narrow-band applications above 2 kMc/s.

Consideration of further improvements in the performance of microwave space-charge valves indicates that further extension of the art will necessitate ambitious assaults on cathode current density and the electron optics of these valves.

6. Acknowledgments

The writer is indebted to the U.S. Army Signal Corps for permission to publish this paper, to Messrs. P. Koskos and R. E. Manfredi for valuable contributions made toward the development of the Z-5267 tetrode and to Messrs. R. P. Watson and H. L. Thorson for stimulating advice and support.

7. References

1. M. T. Vlaardingerbroek, "Small-signal performance and noise properties of microwave triodes", *Philips Res. Repts.*, **15**, pp. 124–221, April 1960.
2. L. J. Giacoletto and H. Johnson, "U.h.f. triode design in terms of operating parameters and electrode spacings" *Proc. Inst. Radio Engrs*, **41**, pp. 51–8, January 1953.
3. J. A. Morton and R. M. Ryder, "Design factors of the Bell Laboratories 1553 triode", *Bell Syst. Tech. J.*, **29**, pp. 496–530, October 1950.
4. H. Rothe and E. Gundert, "Effect of electron transit-time on the efficiency of transmitting tetrodes", *Telefunken-Zeitung*, **25**, pp. 75–82, June 1952.

Manuscript received by the Institution on 18th May 1961 (Paper No. 730).

Radio Engineering Overseas . . .

The following abstracts are taken from Commonwealth, European and Asian journals received by the Institution's Library. Abstracts of papers published in American journals are not included because they are available in many other publications. Members who wish to consult any of the papers quoted should apply to the Librarian, giving full bibliographical details, i.e., title, author, journal and date, of the paper required. All papers are in the language of the country of origin of the journal unless otherwise stated. Translations cannot be supplied. Information on translation services will be found in the Institution publication "Library Services and Technical Information".

TIME DIVIDED SIMULTANEOUS TRANSMITTER-RECEIVER

A communications system called a "Time-divided F.M. Simultaneously Transmitting and Receiving System " in which pulsed f.m. waves of the same carrier frequencies are used for both transmission and reception has been studied by four Japanese engineers. The system's main features are: synchronization is maintained without any special synchronous signals through the use of pulsed f.m. (besides a.m.); the pulsed f.m. waves of minimum band width can always be used by utilizing the "sampling reception" independent of the distance between stations; the use of the modified radio detector results in noise output reduction, so that a squelch circuit is unnecessary.

"A time-divided, simultaneously transmitting-and-receiving system", K. Aoyagi, K. Miyawaki, S. Sogabe and S. Wada. *The Journal of the Institute of Electrical Communication Engineers of Japan*, **44**, pp. 1160–5, August 1961.

A.M.—A.M. TRANSMISSION

A recent Czech paper describes the properties of an a.m.-a.m. system intended for television applications. The properties of the overall signal are analysed and the detection methods are stated. Attention is paid to measurement of the influence of linear frequency distortion on the origin of cross-talk among the transmitted informations. The paper contains the equivalent transmission networks for calculating the cross-talk components.

"The properties and problems of a.m.—a.m. transmission", M. Ptacek. *Slaboproudy Obzor*, **6**, pp. 45–50, January 1962.

REFLECTOR DESIGN FOR RADIO RELAY LINKS

In a recent Czech paper the relations necessary for the design of reflector planes on out-of-sight links are presented in the form of nomograms. There is a nomogram for free space attenuation, between two isotropical radiators, a nomogram for additional attenuation of a single reflector plane, a nomogram for additional attenuation of two reflector planes, and a nomogram for establishing the critical distance between two apertures, with which the transfer equation for free space may yet be considered valid. (See also a previous paper by the author, abstracted in *J. Brit.I.R.E.*, **23**, p. 80, January 1962.)

"Nomograms for the design of plane reflectors for out-of-sight links of radio systems", D. Kupcak. *Slaboproudy Obzor*, **6**, pp. 734–42, December 1961.

DISCONTINUITIES IN COAXIAL CABLES

The variation method has been applied in a recent Polish paper to solve the problem of flat discontinuity in coaxial cables. An integral equation has been deduced for an admittance, fulfilling the conditions of constituent discontinuities E_r and H_y in the aperture and disappearance of E_r in the remaining part of discontinuity plain. This equation is stationary in relation to small changes of E_r and gives the minimum value of Y if the real field arrangement is applied. Then E_r is developed into a full set of cylinder functions. Some special cases of discontinuity in cables are discussed.

"Variation methods of measuring discontinuity in coaxial cables", P. Szulkin. *Rozprawy Elektrotechniczne*, 7, 366–79, 1961.

FADING IN THE U.H.F. RANGE

A recent German paper gives a report on correlation measurements in the receiving end field of scatter links. These measurements were carried out in order to provide some knowledge of the structure and dynamics of the scattering medium. From the results of the correlation measurements the extension of the scatter centres and their own as well as drift velocities are deduced.

"The investigation of fading in the u.h.f. range". J. Grosskopf. *Nachrichtentechnische Zeitschrift*, **14**, pp. 590–604, December 1961.

TURRET TUNERS WITH PRINTED COILS

An Australian paper describes a turret-tuner in which the normal individually replaceable "biscuits" carrying the various coils required for the selection of a particular channel are discarded in favour of wafers with printed coils. As the coils are no longer coaxial, forms of coupling different from the usual mutual coupling are adopted. Some coupling is provided in the stator, but each channel can be individually adjusted at the design stage by arranging to print capacitors or inductors on the wafer. Since the printed coils are of fixed inductance, it is necessary that all stators be alike, and to ensure this, trimming inductors as well as trimming capacitors are included. No provision is made for altering the inductance of the individual oscillator coils, but instead, a mechanical device is used to reset a capacitor in the oscillator circuit. This technique enables the set user to adjust each oscillator to a frequency which is reproduced on switching back to that channel.

"Turret tuners with printed coils", P. T. Rudge. *Proceedings of the Institution of Radio Engineers Australia*, **22**, pp. 748–51, December 1961.